

**CRISPR/Cas9 Genome Editing for the
Generation of YB1 Knockout Human and
Insect Cell Lines for the Production of
Recombinant Adeno-associated Virus Vectors**

by

Kamran Miah

Thesis submitted for the degree of Doctor of Philosophy at the
University of Reading

The School of Biological Sciences

– September 2018 –

Abstract

The use of Adeno-associated viral (AAV) vectors for gene therapy has been most promising because of their safety profile; however, current production methods limit desirable amounts of AAV vectors. Significant efforts have been made to improve AAV production systems, including the optimisation of vector expression cassettes and the regulation of producer cell factors. In this study, we endeavoured to alter the host cell gene expression using CRISPR/Cas9 technology to improve AAV production. We generated knockouts of Y-Box protein (YB)1 in 293T, and the putatively identified homologue *Spodoptera frugiperda* Y-Box protein (SfYB) in Sf9 cells. We present the first example in which CRISPR/Cas9 genome editing can be utilised to regulate cell-intrinsic factors that may be implicated in AAV vector production; although, the disruption of YB1 or SfYB did not generate enhanced AAV vector producer cell lines. YB1 knockout cells presented with heightened sensitivity to chloroquine and limited its use for AAV vector production. The protective function of YB1 to chloroquine-induced cytotoxicity was demonstrated, and analysis of YB1 mutants suggested that its cold shock domain was the principle mediator of this resistance. We also identified YB1 associations with AAV serotype 2 (AAV2) inverted terminal repeat (*ITR*) *in vitro*, and a rather distinct colocalisation between YB1 mutant – encompassing YB1's C-terminal domain (CTD) and AAV2 intact particle and AAV2 capsid to the nucleolar compartment. Therefore, there may be associations between YB1 and its CTD in AAV2 vector production. We also present the advantages of using a repertoire of assays to characterise CRISPR/Cas9-edited cell lines. This included the advantage of establishing clonal populations that were homozygous for their knockout mutation(s), and utilising target-specific antibodies for screening knockouts. Regardless, CRISPR/Cas9 has become a mainstream technology allowing for specific and efficient genome editing, and is revolutionising human gene therapy with the potential of giving rise to an entirely new class of therapeutics.

Original Authorship

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Kamran Miah

Table of Contents

ABSTRACT.....	I
ORIGINAL AUTHORSHIP	II
ACKNOWLEDGEMENTS	VII
LIST OF ABBREVIATIONS.....	VIII
LIST OF FIGURES.....	XIV
LIST OF TABLES	XVIII
CHAPTER 1: INTRODUCTION.....	1
1.1.0 Gene Therapy	1
1.1.1 Definition of gene therapy	1
1.1.2 Conception of gene therapy.....	1
1.1.3 Gene therapy strategies	3
1.1.4 Advantages of AAV vectors in gene therapy	5
1.1.5 Disadvantages of AAV vectors in gene therapy.....	5
1.2.0 Biology of AAV	6
1.2.1 Taxonomy and classification of AAV	6
1.2.2 Structure and organisation of AAV genome.....	7
1.2.3 AAV <i>ITRs</i>	9
1.2.4 Capsid structure	10
1.2.5 AAV serotypes	11
1.2.6 AAV life-cycles and helper-virus functions.....	12
1.2.7 AAV site-specific integration	13
1.2.8 AAV receptor-ligand interaction	15
1.2.9 AAV cell entry	16
1.2.10 AAV trafficking to the nucleus.....	18
1.2.11 AAV DNA replication	21
1.2.12 AAV assembly	23
1.3.0 <i>In vitro</i> production of recombinant AAV vectors	25
1.3.1 Recombinant AAV vectors.....	25
1.3.2 Components required for rAAV production	25
1.3.3 Recombinant AAV vector production.....	26
1.3.4 Advancements in rAAV vector production methods.....	29
1.3.5 Recombinant baculovirus technology	31
1.3.6 Recombinant baculovirus technology for rAAV vector production	33
1.3.7 Identification of human YB1 as a putative AAV restriction factor.....	36
1.4.0 Cellular protein YB1.....	37
1.4.1 Introduction to YB1 and its structure.....	37
1.4.2 YB1's role in transcriptional and translational regulation.....	39
1.4.3 YB1's function in cell proliferation and the cell cycle.....	41
1.4.4 YB1's function in DNA repair	42
1.4.5 YB1's function in cellular stress	44

1.4.6	YB1's function in apoptosis	45
1.4.7	YB1 and viral biology	46
1.5.0	CRISPR and genome editing	48
1.5.1	CRISPR-Cas9 system – evidence of a prokaryotic adaptive immune system	48
1.5.2	Classifications of CRISPR systems in prokaryotes	50
1.5.3	PAM sequence dependency and self-non-self discrimination	52
1.5.4	Mechanism of CRISPR-mediated adaptive immunity	53
1.5.5	Mechanism of spacer acquisition into CRISPR loci	54
1.5.6	Mechanism of crRNA biosynthesis	56
1.5.7	Mechanism of CRISPR interference	58
1.6.0	From adaptive immunity to targeted genome editing <i>in vitro</i> and <i>in vivo</i>	60
1.6.1	Definition of targeted genome editing	60
1.6.2	Features of type II CRISPR-Cas system utilised for genome editing	61
1.6.3	The development of CRISPR/Cas9 system for genome editing and gene manipulation	63
1.6.4	Development of CRISPR/Cas9 system for gene therapy and as a future novel medicine	67
1.7.0	Aims and objectives	75
CHAPTER 2: MATERIALS AND METHODS		77
2.1.0	Materials	77
2.1.1	General laboratory reagents	77
2.1.2	Oligonucleotides (oligos)	78
2.1.3	Chemically competent bacteria – strains	86
2.1.4	Antibodies	86
2.1.5	Eukaryotic cell lines	89
2.2.0	Methods	90
2.2.1	Molecular cloning	90
2.2.2	<i>In silico</i> methods	109
2.2.3	Cell culture, transfections and infections of mammalian and insect cell lines	112
2.2.4	Harvesting genomic DNA and protein	124
2.2.5	Quantification and detection of proteins	126
2.2.6	Specific detection of DNA	132
2.2.7	Pulldown and protein binding assays	135
2.2.8	Statistics	137
CHAPTER 3: ESTABLISHMENT OF YB1 KNOCKOUT CELL LINES FOR RECOMBINANT AAV VECTOR PRODUCTION USING CRISPR/CAS9 GENOME EDITING		138
3.1.0	Introduction	138
3.2.0	Results	140
3.2.1	Cloning <i>YBX1</i> -specific gRNA designs into GeneArt CRISPR nuclease vector	140
3.2.2	Initial screening of transfected 293T cells identified potential YB1 knockout clones	142
3.2.3	Screening YB1 knockout clones for epitope disruption as a result of CRISPR/Cas9 genome editing of <i>YBX1</i>	143
3.2.4	Phenotyping YB1 knockout cell lines using immunocytochemistry (ICC)	145
3.2.5	Sequencing YB1 knockout clones demonstrated mutations by CRISPR/Cas9	148
3.2.6	Alternative splicing is a predicted consequence of CRISPR/Cas9 genome editing identified in YB1 knockout clones	152
3.2.7	Surveyor® mutation screening identifies heterozygous and homozygous genotypes for YB1 knockout clones	161
3.2.8	High resolution melting (HRM) curve analysis distinguishes wildtype alleles from mutated	163

3.2.9	Stable Cas9 expression was not evident in YB1 knockout clones	168
3.2.10	Off-target effects in YB1 knockout clones by Western Blotting.....	170
3.2.11	Cell cytotoxicity phenotype was a consequence of YB1 knockout.....	175
3.2.12	YB1 knockout does not confer an enhanced rAAV2 vector producer cell line	184
3.3.0	Chapter Summary.....	185
CHAPTER 4: THE ROLE OF YB1 ON RECOMBINANT AAV VECTOR PRODUCTION		188
4.1.0	Introduction.....	188
4.2.0	Results	188
4.2.1	Exogenous YB1 expression by plasmid transfection of 293T and YB1 knockout cell lines	188
4.2.2	Infection of YB1 knockout cells with rAAV or lentiviral vectors exhibit comparable cytotoxicity profiles to 293T control	191
4.2.3	Generating stable cell lines expressing YB1 full length and truncation mutants	194
4.2.4	Expression of YB1 and select truncated mutants rescues the cytotoxicity phenotype in YB1 knockout cell lines	207
4.2.5	Examination potential interactions between YB1 and AAV during vector production	215
4.2.6	Impact of YB1 knockout and rAAV2 vector production	229
4.3.0	Chapter summary.....	233
CHAPTER 5: IDENTIFICATION AND MOLECULAR CLONING OF <i>SPODOPTERA FRUGIPERDA</i> Y-BOX PROTEIN HOMOLOGUE (SFYB)		235
5.1.0	Introduction.....	235
5.2.0	Results	236
5.2.1	Prospective Y-Box or YB1 homologue(s) identified in <i>S. frugiperda</i> and <i>T. ni</i>	236
5.2.2	Prospective Y-Box proteins show evolutionary conservation of their CSD	242
5.2.3	Recombinant SfYB(His) ₁₀ expression by transient transfection of Sf9 cells.....	251
5.2.4	Scale-up production of rSfYB(His) ₁₀ was achieved using recombinant baculoviruses.....	259
5.2.5	Purification of rSfYB(His) ₁₀ for the development of antiserum against rSfYB(His) ₁₀	269
5.3.0	Chapter summary.....	274
CHAPTER 6: GENERATION OF SFYB KNOCKOUT CELL LINES FOR RECOMBINANT AAV VECTOR PRODUCTION USING CRISPR/CAS9 GENOME EDITING.....		275
6.1.0	Introduction.....	275
6.2.0	Results	275
6.2.1	Designing and cloning SFYB-specific gRNAs into CRISPR plasmids.....	275
6.2.2	Optimising transfections for targeted knockout of SfYB protein expression	277
6.2.3	Characterising Sf9 cells for SfYB knockout	286
6.2.4	SfYB Y-Box protein homologue associations with AAV	298
6.2.5	The effect of SfYB disruption on rAAV vector production.....	302
6.3.0	Chapter summary.....	310
CHAPTER 7: DISCUSSION AND FUTURE DIRECTIONS		311
7.1	Introduction.....	311
7.2	CRISPR/Cas9 genome editing efficiently generated stable knockout cell lines.....	313
7.3	Complex spectrum of mutation profiles, alternative splicing motifs, and splicing activation are consequences of CRISPR/Cas9 genome editing	314
7.4	Safety profile of YB1 knockout cell lines	316
7.5	Baculovirus technology for <i>S. frugiperda</i> -native recombinant protein expression	318
7.6	SFYB is a <i>S. frugiperda</i> Y-Box protein	320

7.7	Unprecedented consequences to CRISPR/Cas9 genome editing.....	323
7.8	YB1 confers a protective phenotype against chloroquine reagent	330
7.9	Limitations of applying CRISPR/Cas9 genome editing	333
7.10	YB1 and SfYB knockout cell lines do not enhance rAAV vector titres	337
7.11	The role of YB1/SfYB protein and its association with rAAV vectors	339
7.12	Significance of CRISPR/Cas9 genome editing for AAV gene therapy development	344
7.13	Future perspectives.....	346
7.14	Final conclusions	348
BIBLIOGRAPHY.....		350

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Yuan Zhao. Your constant support, motivation, and encouragements have enabled me to be the best I can be, to be the best scientist I can be. You have been patient with me and for that I am especially grateful. Your tutelage over the last few years has been invaluable; I have learned a great deal from your example, and I know I will be able to take on my next challenges with better confidence. To my second supervisor, Prof. Ian Jones, I would like to thank you for inviting and welcoming me to your lab so that I could learn and understand baculovirus technology and apply this into my research focus. I enjoyed my time at the University of Reading as short as it might have been. Thank you also for all the support you have provided me and helping point me in the right direction.

I would also like to extend my gratitude to Stifani Satkunanathan: we were a small group of two, but you were a fantastic labmate during my time at the National Institute of Biological Standards and Control (NIBSC). You made me feel welcome and helped me intergrate into the NIBSC. I would not have come as far as I have without your technical support and training you provided me in various techniques. Next, I would like to say a special thank you to Dr. Giada Matiuzzo and James Ashall for helping me troubleshoot my qPCR woes. This was a rate limiting step for me, and you both helped me overcome this hurdle.

Finally, on a more personal level, I would like to express thank you to my entire family. You have all encouraged me throughout this challenging point in my life. Your unconditional support throughout was truly touching, and I love you all for it. Particularly, I would like to say a special thanks to my brother, Serhan. Your emotional and personal support will not go forgotten. You really have allowed me the space and provided me encouraging words when I needed those most to focus and undertake the challenges thus far. And I know you will continue being the great brother I need in my future endeavours. Thank you all for so much.

List of Abbreviations

%	Percentage
°C	degrees Celsius
A	Amps
A/P	Alanine-proline rich
aa	Amino acid(s)
AAV	Adeno-associated virus
AAVS1	AAV integration site 1
Abs	Antibody
ACE	A/C-rich exon enhancer
AcMNPV	<i>Autographa californica</i> multiple nucleopolyhedrovirus
APS	Ammonium persulphate
BAC	Bacterial artificial chromosome
BEV	Baculovirus expression vector
bp	Base-pair(s)
BR	Basic regions
BSA	Bovine serum albumin
BYB	<i>Bombyx mori</i> Y-Box protein
CaCl₂	Calcium chloride
CAR	Chimeric antigen receptor
Cas	CRISPR-associated protein
Cas9n	Cas9 nickase
Cascade	CRISPR-associated complex for antiviral defence
cDNA	Complementary DNA
CDS	Coding sequence
CLIC	Clathrin-independent carriers
CMV	Cytomegalovirus
co	Codon optimised
CO₂	Carbon dioxide
CQ	Chloroquine
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRa	CRISPR-mediated activation
CRISPRi	CRISPR-mediated interference

crRNA	CRISPR RNA
crRNP	Cas protein-crRNA complex
CRS	Cytoplasmic retention signal
CTCF	Corrected total cell fluorescence
CTD	C-terminal domain
CV	Consensus value
dATP	Deoxyadenosine triphosphate
dCas9	Dead Cas9
dCTP	Deoxycytidine triphosphate
DF	Dilution factor
dGTP	Deoxyguanosine triphosphate
DHFR	Dihydrofolate reductase
DMEM	Dulbecco`s Modified Eagle Media
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
ds	Double-stranded
DTT	Dithiothreitol
dTTP	Deoxyribonucleoside triphosphate
DV	Dengue virus
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EMS	Electrophoretic mobility shift
EMSA	Electrophoretic mobility shift assay
FC	Flow cytometry
FCS	Foetal calf serum
g	Gram(s)
GEEC	GPI-anchored-protein-enriched endosomal compartment
GFP	Green fluorescent protein
gRNA	Guide RNA
h	Hour(s)
H₂O	Water
HAT	Hypoxanthine, aminopterin and thymidine
HCl	Hydrochloric acid

HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGPRT	Hypoxanthine-guanine phosphoribosyl transferase
HIV	Human immunodeficiency virus
HR	Homologous recombination
HRM	High-resolution melting
HRP	Horse-radish peroxidase
HSF	Human Splicing Finder
<i>iap</i>	Alkaline phosphatase
ICC	Immunocytochemistry
IDT	Integrated DNA Technologies
IFN	Interferon
indels	Insertions and/or deletions
ITR	Inverted-terminal repeats
K₂HPO₄	Dipotassium phosphate
kbp	Kilo-base pairs
KCl	Potassium Chloride
kDa	Kilodaltons
ko	Knockout
L	Litre(s)
LB	Luria broth
LTR	Long-terminal repeats
M	Molar
m	milli
mAb	Monoclonal antibody
MgCl₂	Magnesium chloride
MIMIVIRE	Mimivirus virophage resistance element
min(s)	Minute(s)
miRNA	microRNAs
MLV	Murine leukaemia virus
MMTV	Mouse mammary tumour virus
MOI	Multiplicity of infection
mol	moles
MRN	Mre11-Rad50-Nbs

mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein
MSA	Multiples-sequence alignment
MTOC	Microtubule organisation centre
MVM	Minute virus of mice
n	nano
n.d.	Not determined
n/a	Not applicable
N/C	Nuclear:cytoplasmic ratio
Na₂HPO₄	Disodium phosphate
NaCl	Sodium chloride
NaOH	Sodium hydroxide
ncRNA	Non-coding RNA
NEB	New England BioLabs Inc.
NEIL2	Nei-like-2
ng	No growth
NHEJ	Non-homologous recombination
NIBSC	National Institute of Biological Standards and Control
NLS	Nuclear localisation signal
NPC	Nuclear pore complex
nt	Nucleotide(s)
OcMNPV	<i>Orgyia pseudotsugata</i> multiple nucleopolyhedrovirus
OFP	Orange fluorescent protein
oligo(s)	Oligonucleotide(s)
OMIM	Online Mendelian Inheritance of Man
p	pico
pAb	Polyclonal antibody
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PBS-T	PBS plus 0.05% Tween® 20
PCR	Polymerase Chain Reaction
PD	Proportionate distance
PD-1	Programmed cell death protein 1

PEI	Polyethylenimine
pfu	Plaque-forming units
PI3K	Phosphatidylinositol-3 kinase
PIP2	PtdIns(4,5)P2
PLA₂	Phospholipase A2
PML	Promyelocytic leukaemia protein
polh	Polyhedrin promoter
PUMA	p53-upregulated modulator of apoptosis
q	Quantitative
r	Recombinant
Rac1	Ras-related C3 botulinum toxin substrate 1
RBE	Rep-binding element
RIPA	Radioimmunoprecipitation assay
RISC	RNA-induced silencing complex
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
ROI	Region of interest
rpm	Rotations per minute
RT	Room temperature
SAP	Shrimp Alkaline Phosphatase
SaCas9	Cas9 derived from <i>Staphylococcus aureus</i>
Sbjct	Subject
SD	Standard deviation
SDS	Sodium dodecyl sulphate
sec(s)	Second(s)
SeYB	<i>Spodoptera exigua</i> Y-Box protein
Sf-9ET	Sf9 Easy Titre
SFM	Serum-free medium
SfYB	<i>Spodoptera frugiperda</i> Y-Box protein
sgRNA	Single-guide RNA
shRNA	Short-hairpin RNA
SlittoYB	<i>Spodoptera littoralis</i> Y-Box protein
SIYB	<i>Spodoptera litura</i> Y-Box protein

SOC	Super Optimal broth with Catabolite repression
SpCas9	Cas9 derived from <i>Streptococcus pyogenes</i>
SpCas9-HF	SpCas9 high fidelity variant
ss	Single-stranded
SV40	Simian virus 40
T_a	Annealing temperature
TAE	Tris-acetic acid-EDTA
TBE	Tris-boric acid-EDTA
TCID	Tissue culture infectious dose
TCR	T cell receptor
TEMED	N,N,N',N'-Tetramethylethylenediamine
TM	Template modelling
TnYB	<i>Trichoplusia ni</i> Y-Box protein
tracrRNA	Trans-activating crRNA
trs	Terminal resolution site
TSA	Transcriptome Shotgun Assembly
TU	Transducing units
U	Units
UTR	Un-translated region
V	Volts
v/v	Volume over volume
VG	Vector genome
VLP	Virus-like particles
vRNP	Viral RNP
WB	Western blotting
w/v	Weight over volume
YB1	Y-box protein 1
α	Anti
ΔCV	Difference between the wildtype and mutant score
μ	micro

List of Figures

Chapter 1 Figures

Figure 1.1 Genome organisation of AAV2	8
Figure 1.2 AAV2 <i>ITR</i> sequence and secondary structures	10
Figure 1.3 3D-structure of AAV2 capsid.....	11
Figure 1.4 Model for AAV site-specific integration into <i>AAVS1</i> locus.....	15
Figure 1.5 Proposed mechanisms of AAV vector trafficking to the nucleus.....	18
Figure 1.6 Vp1 and Vp2 contain key signals for AAV life-cycle	21
Figure 1.7 Model for AAV genome replication.	23
Figure 1.8 Schematic of a common AAV vector production method by triple transfection	28
Figure 1.9 Recombinant AAV vector production using recombinant baculovirus technology...	36
Figure 1.10 Organisation of YB1 protein.....	38
Figure 1.11 Generic architecture of a CRISPR locus.....	48
Figure 1.12 Three stages of CRISPR-mediated immunity in prokaryotes	54
Figure 1.13 Cas9 mechanism of CRISPR interference of target DNA.....	60
Figure 1.14 CRISPR/Cas9 genome editing of target DNA.....	63

Chapter 2 Figures

Figure 2.1 BLAST/MSA analysis of wildtype vs codon optimised <i>SFYB</i> cDNA sequences.	108
---	-----

Chapter 3 Figures

Figure 3.1 Cloning of <i>YBX1</i> -specific into GeneArt™ linearised, all-in-one, CRISPR nuclease vector for CRISPR/Cas9 genome editing.....	141
Figure 3.2 Establishing YB1 knockout cell lines.....	143
Figure 3.3 Western blot analysis for epitope disruption of selected YB1 knockout cell lines..	145
Figure 3.4 ICC and confocal microscopy analysis for YB1 knockout phenotype.....	147
Figure 3.5 PCR amplification of target sequences from 293T genomic DNA template yielded expected 400-500bp PCR products.....	148
Figure 3.6 Sanger sequencing of YB1 knockout clones A2 and B2 identifies CRISPR/Cas9-mediated mutations.....	150
Figure 3.7 Sanger sequencing of YB1 knockout clone C5 identifies CRISPR/Cas9-mediated mutations.....	151
Figure 3.8 Sanger sequencing identifies stable CRISPR/Cas9-mediated mutations in clones A2, B2, and C5 over a 6 month timecourse	152
Figure 3.9 Predictive analysis for alternative splice variants as a result of 1bp deletion in YB1 knockout clone (A2) <i>YBX1</i> gene	154
Figure 3.10 Predictive analysis for alternative splice variants as a result of 10bp deletion in YB1 knockout clone (B2) <i>YBX1</i> gene	157
Figure 3.11 Predictive analysis for alternative splice variants as a result of A>C substitution and 8bp deletion in YB1 knockout clone (C5) <i>YBX1</i> gene.....	160
Figure 3.12 Surveyor® mutation screening identifies homozygous and heterozygous genotypes by CRISPR/Cas9-genome editing of <i>YBX1</i>	163

Figure 3.13 HRM curve analysis of YB1 knockout clone C5 distinguishes a heterozygous genotype, without wildtype <i>YBX1</i> allele	165
Figure 3.14 HRM curve analysis of sequences upstream of the mutated YBX1 sequence identifies as comparable to 293T control	167
Figure 3.15 HRM curve analysis of sequences downstream of the mutated YBX1 sequence identifies as comparable to 293T control	168
Figure 3.16 Western blot analyses indicate Cas9 expression is transiently expressed after pCRISPR- <i>YBX1</i> sgRNA1-3 transfections.....	170
Figure 3.17 Western blot analysis for off-target effects by CRISPR/Cas9 targeting of <i>YBX1</i>	175
Figure 3.18 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with Calcium Phosphate methodology.....	178
Figure 3.19 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with PEI _{max} reagent.....	180
Figure 3.20 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with FuGENE [®] HD reagent.....	182
Figure 3.21 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with Lipofectamine [®] 2000 reagent	183
Figure 3.22 Recombinant AAV2 vector genome titres compared between 293T and YB1 knockout producer cell lines by qPCR.....	185

Chapter 4 Figures

Figure 4.1 Western blot analysis of exogenous YB1 expression in 293T and YB1 knockout cell lines after pDESTmycYBX1 transfection.....	190
Figure 4.2 LDH cytotoxicity assays for rAAV2GFP transduction of 293T and YB1 knockout cells	192
Figure 4.3 LDH cytotoxicity assays for lentiviral vector (lentiGFP) transduction of 293T and YB1 knockout cells.....	194
Figure 4.4 Designing full length and mutant YB1 coding sequences for cloning into pDUAL vector backbone for lentiviral vector production.....	196
Figure 4.5 Cloning of <i>YBX1</i> full length and mutant YB1 sequences into pDUAL vector backbone for lentiviral vector production.....	198
Figure 4.6 Cloning of <i>YBX1</i> full length and truncated YB1-encoding CDS into pDUAL vector backbone for lentiviral vector production	199
Figure 4.7 Cloning of <i>YBX1</i> full length and truncated YB1-encoding CDS into pDUAL vector backbone for lentiviral vector production	200
Figure 4.8 ICC and laser scanning confocal microscopy of 293T, YB1 knockout cells, and YB1 knockout cells stably expressing mutant YB1 proteins (YB1FL, YB1Δ1-Δ6).....	202
Figure 4.9 Mutant YB1 protein (YB1Δ2, YB1Δ4 and YB1Δ5) show nuclear localisation by line-profiling	205
Figure 4.10 Orthogonal cross-sections further identify YB1Δ2, YB1Δ4, and YB1Δ5 protein localisations.....	206
Figure 4.11 Nuclear/cytoplasmic (N/C) ratios of YB1Δ2, YB1Δ4, and YB1Δ5 compared to 293T control cell line.....	207

Figure 4.12 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h post-treatment with chloroquine	209
Figure 4.13 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h post-transfection using Calcium Phosphate precipitation method supplemented with chloroquine	211
Figure 4.14 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h post-transfection using Calcium Phosphate precipitation method without chloroquine.	213
Figure 4.15 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 72h post-transfection with PEImax.....	214
Figure 4.16 Cloning to generate pAAV2-FLuc plasmid for rAAV2FLuc vector production.....	216
Figure 4.17 ICC for background fluorescence of triple transfected 293T control cells after staining with primary or secondary antibodies	217
Figure 4.18 ICC and confocal laser scanning microscopy analysis of rAAV2 particle co-localisation with YB1 and YB1Δ1-Δ6 truncated mutants.	219
Figure 4.19 ICC and confocal laser scanning microscopy analysis of AAV2 Cap co-localisation with YB1 and YB1Δ1-Δ6	221
Figure 4. 20 ICC and confocal laser scanning microscopy analysis of AAV2 Rep co-localisation with YB1 and YB1Δ1-Δ6	223
Figure 4.21 DNA affinity-YB1 pulldown demonstrates potential YB1-specific binding motifs (GGGGTT and CCTCCT) in AAV- and Adenovirus- associated DNA sequences	227
Figure 4.22 Specific binding of YB1 to putative YB1 binding motifs (GGGGTT and CCTCCT) in AAV2 <i>ITR</i> and <i>rep</i> sequences	229
Figure 4.23 Comparable AAV2 Rep and Vp1-3 expression dynamics between 293T and YB1 knockout cells	230
Figure 4.24 Relative rAAV2 vector genome titres from 293T control, YB1 knockout and YB1Δ1-Δ6 cell lines using qPCR.....	232

Chapter 5 Figures

Figure 5.1 Western blot analysis identifies prospective Y-Box protein homologue in Sf9 cell lines using αYB1	237
Figure 5.2 BLAST screening identified Y-Box protein homologue cDNA sequences for <i>S. frugiperda</i> and <i>T. ni</i>	239
Figure 5.3 Amino acid sequence determined for <i>S. frugiperda</i> Y-Box protein homologue.....	241
Figure 5.4 Amino acid sequence determined for <i>T. ni</i> Y-Box homologue	242
Figure 5.5 Multiple sequence alignment and phylogenetic analyses of human YB1 and Y-Box homologues from <i>B. mori</i> , <i>T. ni</i> , and <i>Spodoptera spp</i> s.....	244
Figure 5.6 Y-Box protein homologues demonstrate conserved intrinsic disorder <i>in silico</i>	246
Figure 5.7 Comparative 3D-structures of YB1 and Y-Box protein homologues' demonstrate structural conservation.....	248
Figure 5.8 Comparative 3D-structures of YB1 and Y-Box protein homologues' CSDs demonstrate structural homology to PDB templates.....	251
Figure 5.9 Cloning of <i>SFYBco</i> transgene into pEx™-1 cloning vector	253
Figure 5.10 Sub-cloning of <i>EGFP</i> transgene into pEx™-1 cloning vector	255
Figure 5.11 Screening for His-tagged rSfYB(His) ₁₀ expression after plasmid transfection of Sf9 cells	257

Figure 5.12 Recombinant SfYB(His) ₁₀ expression and pulldown after plasmid transfection of Sf9 cells.....	259
Figure 5.13 Cloning by In-Fusion™ of <i>His-SFYBco</i> transgene into pTriEx™-1.1 baculovirus cloning vector.....	261
Figure 5.14 Recombinant SfYB(His) ₁₀ expression confirmed after BacSFYB(His) ₁₀ infections ..	263
Figure 5.15 Intracellular staining for His-tag protein may not correlate expression levels for target rSfYB(His) ₁₀	265
Figure 5.16 Optimisation of rSfYB(His) ₁₀ expression for purification.....	267
Figure 5.17 Recombinant SfYB(His) ₁₀ expression localised in the membrane fraction of infected Sf9	269
Figure 5.18 Purification of His-tagged rSfYB(His) ₁₀ protein	271
Figure 5.19 Antiserum raised against rSfYB(His) ₁₀ shows specificity for SfYB	273

Chapter 6 Figures

Figure 6.1 Inferring <i>SFYB</i> gene exon and intron sequences using <i>BYB</i> and <i>SLYB</i> genes as genomic scaffolds.....	276
Figure 6.2 Cloning of gRNAsf1-3 oligos into GeneArt™ linear, all-in-one nuclease vector.....	277
Figure 6.3 Transfection of Sf9 cells with pCRISPR- <i>SFYB</i> sgRNAsf1-3	279
Figure 6.4 Optimisation of Sf9 transfections using CRISPRmax™ Reagent and different diluents by flow cytometry analysis.....	281
Figure 6.5 IVT reactions and purification of sgRNAs for <i>SFYB</i> targeting.....	282
Figure 6.6 Strategy employed to transfect (multi-run transfection) Sf9 cells with sgRNA and recombinant Cas9 complex for targeted disruption of <i>SFYB</i> gene	286
Figure 6.7 Establishing SfYB knockout cell lines.....	287
Figure 6.8 Screening multi-run transfected Sf9 cells with sgRNAsf1 for SfYB disruption	289
Figure 6.9 Confirmation of partial knockout and knockout phenotypes using SfYB-specific antiserum	290
Figure 6.10 Screening for SfYB knockout cell lines transfected with sgRNAsf2 for <i>SFYB</i> disruption.....	291
Figure 6.11 Screening for SfYB knockout cell lines transfected with sgRNAsf3 for <i>SFYB</i> disruption.....	292
Figure 6.12 Growth defect of SfYB knockout cell line compared to Sf9	293
Figure 6.13 Characterising SfYB knockout cell lines using ICC and confocal laser scanning microscopy	295
Figure 6.14 Genomic profiling of CRISPR/Cas9-mediated genome editing of the <i>SFYB</i> gene ..	297
Figure 6.15 SfYB binding to DNA demonstrates shift in electrophoretic mobility by agarose gel electrophoresis	299
Figure 6.16 ICC and confocal laser scanning microscopy of SfYB and AAV2 Rep, Cap, and particle localisations.....	301
Figure 6.17 SfYB disruption or exogenous expression does not affect the efficiency of baculovirus infection.....	302
Figure 6.18 LDH cytotoxicity assays for baculovirus infection of Sf9 and SfYB knockout cell lines	303
Figure 6.19 Gene transfer assays of 293T with crude rAAV -2 or -8 vectors	305

Figure 6.20 SfYB disruption did not influence AAV Rep or Cap (Vp1-3) protein expression	307
Figure 6.21 Comparable rAAV VG titres were detected between parental Sf9 and Sf9ex1.7/10X cell lines.....	309

List of Tables

Table 1.1. List of human clinical trials harnessing zinc-finger nuclease to target HIV1 infection	70
Table 1.2. List of human clinical trials utilising CRISPR/Cas9 genome editing.....	74
Table 2.1 General reagents used	77
Table 2.2 List of cloning and sequencing oligos.....	79
Table 2.3 <i>YBX1</i> -specific gRNA designs to target <i>YBX1</i> gene	81
Table 2.4 <i>SFYB</i> -specific gRNA designs to target <i>SFYB</i> gene	82
Table 2.5 <i>SFYB</i> -specific gRNA designs to target <i>SFYB</i> gene	83
Table 2.6 List of oligos for DNA-affinity YB1 pulldown	84
Table 2.7 List of primary antibodies.....	87
Table 2.8 List of secondary antibodies.....	89
Table 2.9 List and details of cell lines.....	89
Table 2.10 <i>wildtype SFYB and SFYBco</i> gene string sequence.....	107
Table 3.1 Predictive assessment for altered splicing motifs after modelling clone B2's mutated sequence into HSF v3.0.....	156
Table 3.2 Predictive assessment for altered splicing motifs after modelling clone C5's mutated sequence into HSF v3.0.....	159
Table 3.3 Potential off-target sites identified for gRNA1 targeting of <i>YBX1</i> gene.....	172
Table 3.4 Potential off-target sites identified for gRNA2 targeting of <i>YBX1</i> gene.....	173
Table 3.5 Potential off-target sites identified for gRNA3 targeting of <i>YBX1</i> gene.....	174
Table 4.1 List of YB1FL and YB1 mutants and their subcellular localisation and characteristics	203
Table 4.2 YB1 binding motifs identified in AAV and Adenovirus sequences	225
Table 5.1 List of Y-Box protein homologues explored in the present study.....	243
Table 5.2 Y-Box protein model templates and TM-scores using IntFOLD and TM-align	250
Table 6.1 Observations from single-cell cloning of Sf9 cells 72h post-transfection with sgRNAs and recombinant Cas9	284
Table 6.2 Observations from single-cell cloning of Sf9 cells with various Sf900™ II SFM variations.....	285

Chapter 1: Introduction

1.1.0 Gene Therapy

1.1.1 Definition of gene therapy

Gene therapy is defined as the directed and targeted delivery of exogenous genetic material to correct a negative phenotype or disease. The goal of gene therapy is to cure the clinical status of patients, whom suffer from certain heritable or acquired genetic diseases. In this way the recipient of gene therapy is corrected of their genetic defect either by mediating the repair of patients' genetic information (Urnov *et al.*, 2005; Wu *et al.*, 2015), or inserting the functioning gene into the target tissue(s) (Anderson, 1984).

1.1.2 Conception of gene therapy

The core principles behind gene therapy derived from initial studies that propagated within the last century with work performed on bacteria. Frederick Griffith examined the then unknown phenomenon of transformation (Griffith, 1928) by mixing live, non-virulent R form of Type I pneumococcus with heat-inactivated virulent S strain. Subsequent infection of this mixture in mice developed active infection, pneumonia, and death. Additionally, viable isolates of S form of Type II pneumococcus were cultivated from infected mice. Thus Griffith concluded the R form had converted to the more virulent micro-organism. Studies that followed, shortly thereafter, showed that the key factor that permitted such transformation was mediated by the transfer of deoxyribonucleic acid (DNA) (Avery *et al.*, 1944).

Additional gene transfer mechanisms were shortly discovered, and perhaps an important mediator of horizontal transfer of genetic material was via bacteriophages, in a process dubbed 'transduction' (Zinder and Lederberg, 1952). These particulates invade and inject their genetic material into target bacterial cells, which then facilitates their replication and consequently induces cell lysis. Or a dormant infection establishes with integration of bacteriophage genetic material into the target genome. Upon particle assembly, bacteriophages are capable of acquiring the target's genetic information. Subsequent infection facilitates the exchange of genetic material between bacteria (Zinder and Lederberg, 1952; Wilson *et al.*, 1979). This process contributes to the rapid acquisition of antibiotic resistance in certain bacteria (Colomer-Lluch *et al.*, 2011). The process exemplifies the extent at which

bacteriophages are capable of functioning as transforming mediators.

Further to this, work performed by Waclaw Szybalski demonstrated the potential for rescuing genetic defects by the acquisition of functional exogenous DNA (Szybalska and Szybalski, 1962). Cells require the synthesis of nucleic acids for growth and survival. This synthesis is programmed in the form of dihydrofolate reductase (DHFR), and when inhibited, an alternative pathway is utilised in the form of hypoxanthine-guanine phosphoribosyl transferase (HGPRT). Taking advantage of this, Szybalska generated HGPRT⁺ and HGPRT⁻ clones from the human bone marrow cell line, D98S. Inhibition of DHFR by aminopterin (supplemented as hypoxanthine, aminopterin, and thymidine [HAT] medium) promoted the survival of HGPRT⁺ derivatives only because the alternative HGPRT-mediated synthesis of nucleic acid compensated for the lack of DHFR activity. Subsequent isolation of genomic DNA from the HGPRT⁺ cells and transformation of HGPRT⁻ cells with said isolated DNA resulted in the complete rescue of HGPRT⁻ cells when grown in HAT medium. Furthermore, daughter cells of transformed cells also showed a similar phenotype to the HGPRT⁺ derivative, indicating stable transfer of the rescued gene (Szybalska and Szybalski, 1962). This represented one of the earliest indicators that the physical transfer of genetic information could be a means to produce a desired phenotype.

The phenomena of transduction, genetic transfer, and the concept of rescuing genetic defects fuelled research innovation with extension to eukaryotic viruses. One driver of this potential application was the observation that pseudovirions were an additional product of polyomavirus (Winocour, 1968) or simian virus 40 (SV40) (Trilling and Axelrod, 1970) infections *in vitro*. Pseudovirions refer to progeny particles that harbour packaged fragments of host DNA instead of virus-specific genome. Observations made by Osterman *et al.* (1970) showed that pseudovirions were capable of intracellular uncoating and, therefore, exogenously-acquired DNA could be shuttled safely into infected cells. Although, no indication of the fate of the shuttled host DNA fragments could be deduced from initial studies, the aforementioned work provided the initial premise that genetic information could be delivered safely *in vivo*. It quickly became apparent that the mammalian SV40 virus could be manipulated to function as a transfer vector of genetic information into eukaryotic cells, and propagate recombinant vectors from cultured mammalian cells (Ganem *et al.*, 1976; Goff and Berg, 1976; Nussbaum *et al.*, 1976). Replacing SV40 genome sequences, namely the sequences encoding the SV40 large T antigen and/or sequences encoding the structural proteins Vp1-3, with a desired transgene

permitted *in vitro* production of replication-defective recombinant SV40 vectors in COS-7 cells (Strayer, 1996). Vector production was permissive because COS-7 cells provided the large T antigen *in trans* (Strayer, 1996). The therapeutic potential of recombinant SV40 vectors has been demonstrated in animal models of select human diseases (Goldstein *et al.*, 2002; Duan *et al.*, 2004; Vera *et al.*, 2007). However, a considerable limitation of recombinant SV40 vector technology and its use in clinical studies included the accumulation of replication-competent SV40 vectors during production processes (Vera *et al.*, 2004). This was likely as a result of recombination between recombinant SV40 DNA (for vector production) in COS-7 cells that were originally immortalised by the introduction of origin of replication-mutant SV40 viral DNA (Gluzman, 1981).

1.1.3 Gene therapy strategies

It became essential to identify more appropriate viral vectors for the delivery of exogenous genetic material into target cells, especially those that could stably transduce target cells for therapeutic purposes. Additional studies revealed the capacity for gene transfer by means of alternative viral vectors, including: lentivirus, Adenovirus and Adeno-associated viruses (AAVs) (Volpers and Kochanek, 2004; Kotterman and Schaffer, 2014). Therefore, the principles of gene therapy involved introducing nucleic acid to be administered directly to humans with the aim of genetic engineering of target cells. This would allow substitution or replacement of defective gene sequences, and potentially provide long-term curative benefit to those patients who suffer from acquired or genetic disease.

The predicted collective of monogenic diseases affect a great number of people worldwide. A list of approximately 10,000 monogenic diseases is available in the Online Mendelian Inheritance in man (OMIM) database. The prospect of alleviating these conditions by a single-administration, curative strategy is highly attractive. And with enhanced next-generation sequencing technologies, the precise genes or genetic defects responsible for a great many genetic disorders (approximately 50%) have been identified in good confidence (Boycott *et al.*, 2013). With enhanced molecular understanding of genetic diseases, the targeting efficiency of these identified gene defects can be improved substantially with greater overall clinical success.

To tackle this viral vector-based gene therapy platforms have been actively developed. Their applications have been widely extended from the original concept of reintroducing the correct

gene to target cells or tissue. Such includes direct correction of the defective genes using genome editing technologies (this will be briefly discussed in section 1.6.4). However, it is worth mentioning that alternatives to viral vector based gene therapies are actively being researched and have even undergone clinical trials. The intention is to expand gene therapy platforms to overcome limitations of viral vector and/or rAAV vector strategies for gene therapy. Such includes non-viral delivery methods that involve the gene transfer of therapeutic cDNA-encoding gene, plasmid, or RNA. Non-viral vector gene transfer methods involving electroporation (Neumann *et al.*, 1982) of desired nucleic acids have shown some clinical translation potential, either *in vivo* and/or *ex vivo* (Brown *et al.*, 2009). However, given the nature of the gene transfer technique, the use of electric fields and impulses can incur localised cell death and damage, which is more apparent than physical gene transfer methods such as viral vector strategies (Lefevre *et al.*, 2002; Kubota *et al.*, 2005).

Additional non-viral vector strategies can also involve introducing nucleic acids complexed to liposomes (for example pGM169/GL67A), polymers, or gold nanoparticles for gene therapy applications (Alton *et al.*, 2015; Lee *et al.*, 2017b; Sun *et al.*, 2017). The use of a viral vector to shuttle the desired nucleic acid is omitted. Many advantages are proposed for using non-viral gene delivery methods over viral delivery methods. This includes the increased safety from the naked DNA randomly integrating into the host genome and potentially encouraging insertional mutagenesis. The cDNA/RNA used in non-viral delivery methods is largely non-integrating into host genomes, and are unlikely to affect the genomics of proto-oncogenes or tumour suppressor genes. Additionally, non-viral methods are associated with low immunotoxicity given the absence of viral proteins, which reduces the chance for adverse immunological responses to the gene delivery as compared to using viral vector methods (Chen *et al.*, 2010; Kelley *et al.*, 2018). This is largely because recombinant viral vectors are typically derived from naturally occurring viruses, from which a subset of the human population may be experienced against and already have an immunological memory component ready to mount exaggerated immune responses against cells targeted by viral vectors (Veron *et al.*, 2012). Finally, viral vectors often come with a packaging limit for the desired therapeutic gene as is seen with rAAV vectors (up to 4.5kbp). The restriction in cDNA size is in theory lifted using non-viral methods as packaging capacities do not need to be considered. Having said this, gene delivery efficiencies are compromised with significantly large cDNAs regardless of complexing with liposomes, polymers, or gold particles (Ribeiro *et al.*, 2012).

1.1.4 Advantages of AAV vectors in gene therapy

To facilitate the above, vector-based gene therapy has become an extremely attractive approach. AAV vectors show considerable promise as a system to deliver recombinant nucleic acid into target cells. The great wealth of information available on AAVs and its biology has permitted its easy manipulation. Knowledge in AAV capsid structure and design has enabled commercial applications for research and translational benefit. Wildtype AAVs have shown, until more recently (and briefly discussed in Section 1.1.5), no known association with active human disease or cytopathogenicity, and thus potentiates the viral system as safe (Nathwani *et al.*, 2007; van Gestel *et al.*, 2014). Moreover, AAV transduces a broad range of cells including non-dividing cells. This poses a complication and limitation when concerning other vector-based strategies, which have a marked preference for transducing actively dividing cells only. However, given AAV's tractable infectivity towards either cell states (dividing or non-dividing), vectors derived from this system are able to circumvent this limitation, and broaden cell targeting capabilities. By extension, the prospect of targeting a greater repertoire of genetic diseases is thus more permissible.

It should be appreciated, that unlike SV40-based gene transfer, use of recombinant (r)AAV vectors is associated with long-term and stable transgene expression (Kaspar *et al.*, 2005) – a prerequisite for successful gene therapy. Recombinant AAV has therefore been used in gene therapy clinical trials for a number of diseases, such as: Duchenne muscular dystrophy (Bowles *et al.*, 2012), haemophilia B (Nathwani *et al.*, 2011; Nathwani *et al.*, 2014; French *et al.*, 2018; Miesbach *et al.*, 2018), and cystic fibrosis (Moss *et al.*, 2004). More promising is the fact that the first EU-licensed gene therapy product, Glybera, and the even more recent US-approved Luxturna, are both rAAV-based gene therapy products for the treatment of hereditary lipoprotein lipase deficiency (Carpentier *et al.*, 2012) and Leber's congenital amaurosis (Bennett *et al.*, 2016), respectively. Collectively, these examples demonstrate that rAAV-based gene therapies are a promising advancement as a modern healthcare and medicine.

1.1.5 Disadvantages of AAV vectors in gene therapy

Despite the advantages of using rAAV vectors for gene therapy applications (as briefly outlined above and in further detail in the rest of the literature review), the vector system is not without some drawbacks that limit its extended use as a gene therapy platform. Firstly, the packaging capacity of rAAV vectors is very limiting, where wildtype AAV genomes are approximately 4.7kb in length (Srivastava *et al.*, 1983). Therefore, after removal of the *rep* and

cap encoding sequences, little cargo space is left for the desired therapeutic transgene. An attempt to overload rAAV vectors is met with challenges in packaging intact vector genomes – vector genomes are found truncated at the 5'-end and/or showed impaired transduction efficiencies *in vitro* (Wu *et al.*, 2010). Next, rAAV vector genomes chiefly persists episomally in the nucleus of transduced cells (Penaud-Budloo *et al.*, 2008), and rarely integrates randomly into the host genome (Inagaki *et al.*, 2008). Therefore, depending on the target cell type, episomal rAAV vector genomes can be simply diluted out by cell division, leading to a transient state of gene correction. Additionally, given that in most instances the capsids of naturally occurring AAV serotypes are used to pseudotype rAAV2 vector genomes, of which wildtype counterparts naturally infect humans, and consequently contribute to an immune response and immune memory to rAAV vectors. Ultimately, the host immunological response to rAAV vectors is evident (Veron *et al.*, 2012), and restricts the efficacy of rAAV vectors as a gene therapy tool (Janelidze *et al.*, 2014).

Furthermore, the safety profile of using rAAV vectors for gene therapy has become a controversial topic despite the number of clinical trials to date that suggest its general safety profile. Potential genotoxicity by rAAV vectors was first implied in murine models that demonstrated a significant risk in hepatocellular carcinoma development (Donsante *et al.*, 2001; Donsante *et al.*, 2007). The genotoxicity issue was further raised by Nault *et al.* (2015), of which group identified the integration of wildtype AAV2 genome sequences in 11/193 human hepatocellular carcinoma samples. The integration occurred in key genes known to be cancer drivers and correlated with their overexpression (CCNA2, TERT, CCNE1) (Nault *et al.*, 2015). Having said this, no long term studies on large animal models thus far have demonstrated rAAV-mediated genotoxicity, post-transduction with rAAV vectors in a clinical context (Gil-Farina *et al.*, 2016). Nonetheless, this potential revelation that AAV, and by extension rAAV, may show some tendency for insertional mutagenesis and hepatocellular carcinoma risk, may warrant long term followup and observation of clinical trial participants for tumour formation. This would likely include identifying AAV vector integration sites to evaluate the risk of insertional mutagenesis in studies involving rAAV vector transduction.

1.2.0 Biology of AAV

1.2.1 Taxonomy and classification of AAV

Human AAV was first identified as contaminants of simian Adenovirus preparations (Atchison *et al.*, 1965). AAVs are small (20-25nm), non-enveloped, DNA viruses which have been classed

into the *Parvoviridae* family. *Parvoviridae* family of viruses are among the smallest viruses known, and comprise a wide number of non-enveloped viruses with capsid shells that form an icosahedral structure of T=1 symmetry (Xie *et al.*, 2002). AAVs belong to its own genus, Dependovirus, originally described due to their tendency to require helper-viral functions to propagate. Therefore, AAVs are non-productive alone, and AAVs' life-cycle is incomplete unless aided by the presence of an unrelated DNA virus – namely Adenovirus (Hoggan *et al.*, 1966) or Herpes Simplex Virus (HSV) (Buller *et al.*, 1981). Upon super-infection these helper viruses act to change the intracellular milieu and permit AAV gene expression and replication. Adenovirus provides AAV with early Adenoviral proteins: E1A, E1B, E4 and E2A. These helper proteins function to promote host-cell entry into S-phase and DNA replication (Samulski and Shenk, 1988). It also has been described that AAV replication can be stimulated by cellular genotoxic stresses (experimentally simulated by ultra-violet light, chemical carcinogens, or inhibitors of DNA replication) in the absence of helper virus, resulting in infectious AAV progeny (Yalkinoglu *et al.*, 1988). In contrast to this, when helper functions are absent, AAV establishes latent infection with site-specific integration into chromosome 19q13.4 (Kotin *et al.*, 1990). Thirteen serotypes of AAV (that are capable of infecting humans and primates) have been described; with AAV2 being the best characterised serotype. The remainder of this literature review will focus on this particular serotype as a prototype for the entire family.

1.2.2 Structure and organisation of AAV genome

All AAV serotypes package a single-stranded (ss)DNA genome of either polarity (plus or minus strands) at equal efficiency during assembly of AAV progeny (Steinbach *et al.*, 1997). AAV DNA genome is approximately 4.7kb in length (Srivastava *et al.*, 1983), and at either end (5' and 3' ends) are specialised T-shaped hairpin secondary structures. These form as a result of 145 nucleotide (nt) sequences called inverted terminal repeats (*ITRs*), and flank AAV coding regions.

The organisation of AAV genome's coding sequence can be summarised by the possession of two main coding gene cassettes or open reading frames (ORFs) – *rep* and *cap* (Fig. 1.1). The primary ORFs contain a region in which the *rep* and *cap* genes are overlapping by a small intron sequence. The *rep* and *cap* encode for four non-structural proteins and three structural proteins, respectively. Although, an additional non-structural protein involved in AAV assembly (assembly-activating protein, AAP) has been identified by Sonntag *et al.* (2010), and is encoded within the *cap* ORF. More recently, an additional AAV2 gene has been

characterised (AAV2 X). The AAV2 X gene was found in the more 3'-end of the AAV2 genome, which is associated with its own promoter – p81 (Cao *et al.*, 2014). It was found that the product of the X gene enhanced autonomous and helper-directed DNA replication six-fold (Cao *et al.*, 2014).

A total of three major messenger (m)RNA transcripts are generated from the two ORFs of AAV genome's plus strand (Jay *et al.*, 1979). Transcription initiation of each transcript is controlled by individual promoters termed p5, p19 and p40 (defined by their relative location within the genome). Both ORFs share a common 3'-end polyadenylation site to dictate transcriptional termination. Moreover, through the use of multiple promoters and hijacking of the host cell transcriptional machinery a total of eight proteins are easily encoded within the short coding sequence of AAV.

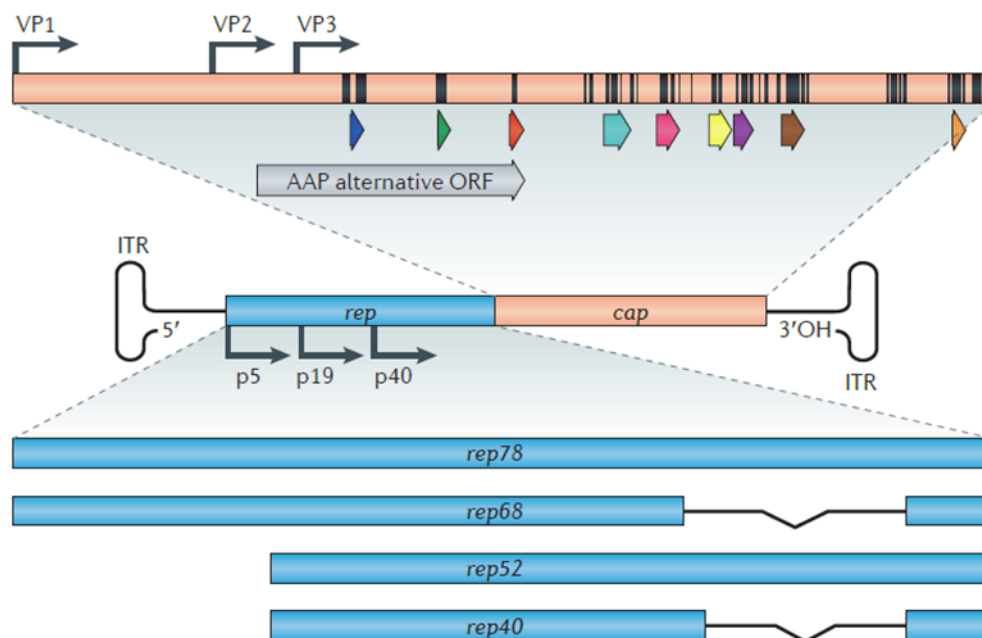


Figure 1.1 Genome organisation of AAV2. Representative organisation of AAV2 genome, with *rep* and *cap* ORFs flanked by *ITR* sequences. Further dissection of *rep* and *cap* ORFs is also shown, with several main transcripts encoded by two ORFs – coding for three structural proteins (Vp1-3), and four non-structural proteins (Rep78, -68, -52 and -40). Black lines within *cap* represent surface-exposed aa residues. AAP alternative reading frame is denoted by the grey arrow. Coloured arrows indicate separate hypervariable regions. Figure from Kotterman and Schaffer (2014).

Because of the limited coding capacity of AAV, AAV is restricted in its self-sufficiency. For instance, AAV does not encode its own RNA polymerase for transcription. To compensate, AAV is dependent on the host cell machinery to engage transcriptional activity and processing. Nonetheless, through transcriptional processing of alternate splice variants, the *rep* ORF encodes four non-structural proteins termed Rep78, Rep68, Rep52 and Rep40, which engage

in a number of processes – namely replication. Expression of Rep78 and Rep68 are under the transcriptional control of p5 promoter, whereas Rep52 and Rep40 are determined by the p19 promoter. Rep68 and Rep40 are C-terminal truncated splice variants of the Rep78 and Rep52 major transcripts, respectively. The *cap* ORF, instead, is under the transcriptional control of the p40 promoter and encodes the capsid structural proteins Vp1, Vp2 and Vp3, as well as AAP. Vp1-3 makes up the core components of AAV's protein shell, and also determines the AAV serotype. The major transcripts undergo splicing events due to the presence of the small intron sequence found within the *rep-cap* overlapping region (Qiu *et al.*, 2003). The small intron possesses relevant splice donor and splice acceptor sites to facilitate splicing of either ORF, and therefore contributes to the repertoire of proteins expressed by wildtype AAV genome.

1.2.3 AAV ITRs

The 5' and 3' ends are described as the *ITRs* of AAV genome, where the terminal 145nt form T-shaped hairpin secondary structures. The initial-most 125nt form imperfect palindromes that folds upon itself to adopt the T-shaped hairpins. The *ITR* is a key requirement for certain aspects of AAV biology; for example a key role of the *ITR* is to facilitate AAV genome replication by creating double-stranded (ds)DNA duplex that helps self-prime DNA synthesis. In addition to this the *ITRs* also possess *cis*-elements such as the Rep-binding elements (RBE and RBE'), and terminal resolution site (*trs*) involved in the nicking of dsDNA after completion of DNA replication. These are all necessary for active replication of AAV genome and packaging of genomes into preformed AAV capsids, to then form infectious units. Wildtype AAV is known to integrate site-specifically into the infected cell's genome into chromosome 19q13.4, and this process is largely dependent on the *ITRs*. A representative diagram of AAV2 *ITR* is portrayed in Fig. 1.2.

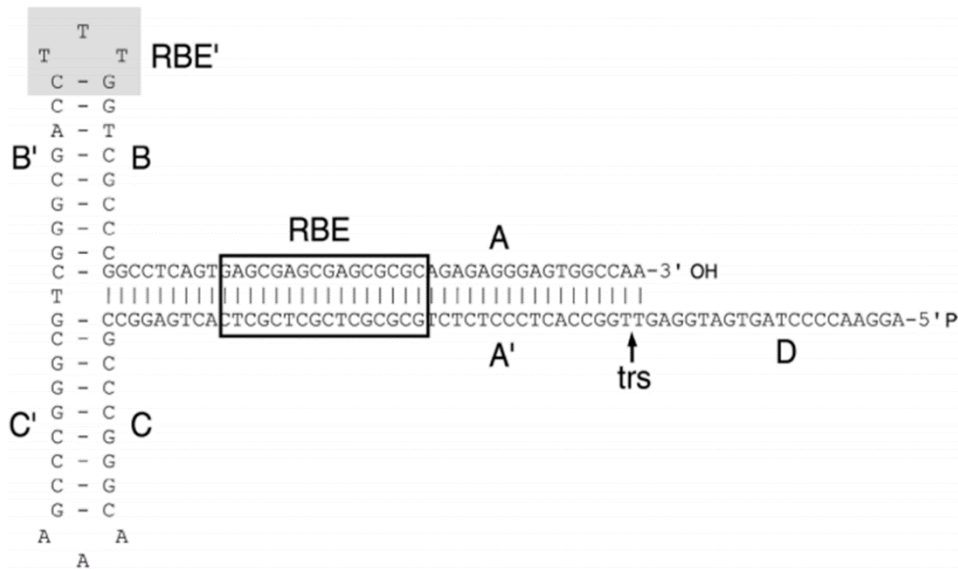


Figure 1.2 AAV2 ITR sequence and secondary structures. Secondary structure of the left ITR is depicted, with annotations describing relevant features of the AAV2 ITR. AAV ITR is composed of two palindrome sequences (B-B' and C-C') that make up the arms of the T-hairpin structure, and the longer palindrome sequence (A-A') that makes up the stem of the ITR. The D-sequence on either end of the AAV genome is the point where the AAV genome becomes single-stranded. RBE, Rep binding element; RBE', second Rep binding element; *trs*, terminal resolution site. Figure from Goncalves (2005).

1.2.4 Capsid structure

As depicted previously the *cap* ORF encodes the capsid proteins (Vp1-3). Each share C-terminus residues; however, Vp2 and Vp3 are N-terminus truncated splice variants of Vp1. Vp1-3 are 87kDa, 73kDa and 61kDa in size, respectively. The capsid of AAV virions is composed of sixty subunits made up of these three core proteins. The triplets of proteins arrange themselves in a specific 1:1:10 ratio of Vp1, Vp2 and Vp3, respectively, to form individual subunits (Agbandje-McKenna and Kleinschmidt, 2011). Altogether these 60 subunits assemble to form an icosahedral structure with T=1 icosahedral symmetry (Fig. 1.3A) as determined by X-ray crystallography at 3Å resolution for AAV2 (Xie *et al.*, 2002). The 3D-structure of several AAV serotypes has also been determined by X-ray crystallography, including: AAV4 (Padron *et al.*, 2005) and AAV5 (Walters *et al.*, 2004). Shared characteristics are evident between AAV serotype capsids, and these shared topological features are mainly found at each axis of symmetry. For instance, the most prominent features of AAV capsids are protrusions that appear at the three-fold symmetry axis, and a cylindrical channel that appears at the five-fold symmetry axis (Fig. 1.3B). Fig. 1.3B also depicts prominent depressions at the two-fold symmetry axis, which are immediately followed by the protrusions at the three-fold symmetry axis (O'Donnell *et al.*, 2009).

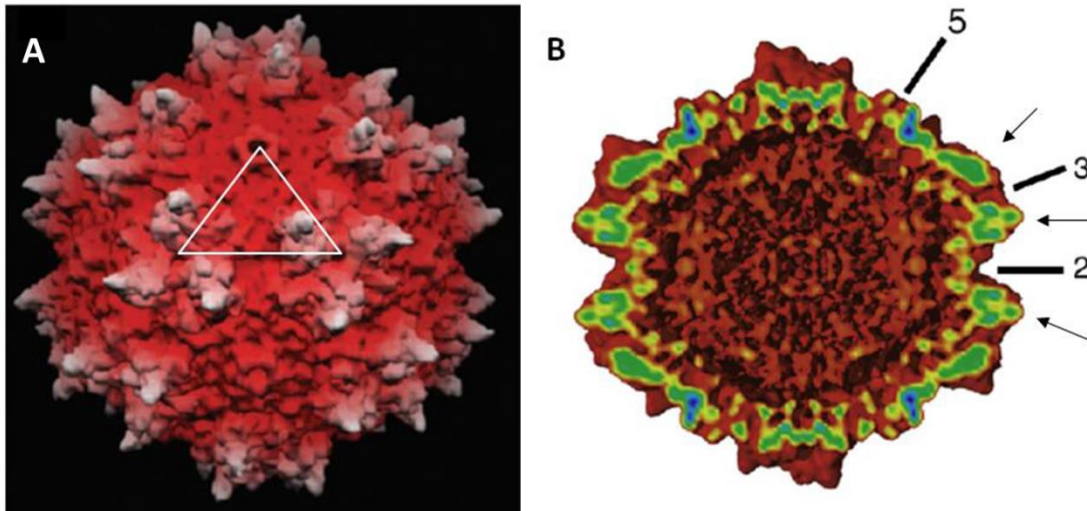


Figure 1.3 3D-structure of AAV2 capsid. **A)** 3D representation of AAV2 capsid shell, where the white triangle represents one subunit composed of the Vp1-3 at 1:1:10 stoichiometry. These assemble to form the icosahedral structure. Topology is coloured to reflect distance from centre of particle – white, furthest distance from centre, and red is closest to centre. Figure adapted from Xie *et al.* (2002). **B)** Cross-sectional of AAV2 capsid with symmetry axes 2, 3, and 5 annotated. Arrows indicate protrusions at the 3-fold symmetry axis that flank 2-fold symmetry axis depression. Figure adapted from O'Donnell *et al.* (2009).

1.2.5 AAV serotypes

Serotypes are defined as an isolated virus that does not cross-react with neutralising antibodies or antiserum that is originally specific against other existing forms. AAV2 was the first isolated serotype to be cloned into a bacterial plasmid by Samulski *et al.* (1982). Therefore, it became the key serotype used for clinical and research purposes. Consequently, AAV2 biology is best characterised and extensively elucidated in comparison to all other naturally occurring AAV serotypes.

The initial most AAV serotypes 1-4 and 6 were mainly isolated as contaminants of Adenovirus preparations (Atchison *et al.*, 1965), whereas AAV5 was identified from human penile genital warts (Bantel-Schaal and zur Hausen, 1984). Several more recent AAV serotypes have also been identified, including AAV7-9, as well as well over 100 AAV variants (Gao *et al.*, 2002a; Gao *et al.*, 2004; Mori *et al.*, 2004; Schmidt *et al.*, 2006). This novel repertoire of AAV serotypes was identified through PCR-based strategies that targeted homologous regions of the *cap* gene in a number of human tissues (Gao *et al.*, 2004). It then became apparent that some AAV serotypes showed distinct *in vivo* transduction properties despite significant homology at the DNA-level. For example, isolation of AAV8 showed remarkable transduction into murine liver compared to AAV2 (Sands, 2011). The distinct transduction biology between serotypes encouraged expansion of AAV as a tool for gene therapy. Where the select tropism

displayed by particular serotypes proves advantageous in a clinical setting by avoiding broad, non-specific transduction, and thus allowing for more site-specific targeting of cell or tissues.

1.2.6 AAV life-cycles and helper-virus functions

Upon infection into target host cell, AAV typically undergoes one of two life-cycles – lytic or lysogenic. Lytic infection predominates when helper functions and virus are made available. Productive AAV results and is characterised by active genome replication, gene expression of viral-encoded genes, and the eventual production of AAV virions. Helper virus-mediated lysis of host cell results in the release of AAV virus progeny.

The repertoire of Adenovirus genes that mediate the helper functions for AAV production were identified as *E1a*, *E1b*, *E2a*, *E4orf6* and virus-associated RNAs (*VA RNAs*) (Samulski and Shenk, 1988). The encoded proteins and RNAs permit AAV production by affecting different aspects of AAV life-cycle as well as the biology of the infected host cell. To start with, *E1a* encodes E1A protein, which up-regulates cyclin E and cdc25A expression and activity and promotes host cell entry into active S-phase of the cell cycle (Spitkovsky *et al.*, 1996). Additionally, E1A is directly implicated in the pRb/E2F-1 pathway, by binding to the pRb family of proteins to release key transcription factors that are defined as key regulators of S-phase entry from the G1 stage of the cell cycle (Nevins, 1990). Ultimately there is gross manipulation of the infected cell that prompts cell cycle entry and up-regulation of DNA replication and synthesis machineries. E1A additionally functions to regulate virus replication by promoting transcriptional up-regulation of the AAV promoters, p5 and p19 (Tratschin *et al.*, 1984). E1A is capable of stabilising p53 – the tumour suppressor gene product, to elicit pro-apoptotic functions. This is mediated through E1A's ability to render p53 incapable of proteasomal degradation (Lowe and Ruley, 1993; Li *et al.*, 2004). Although, remaining helper functions have evolved to prevent E1A-mediated p53-induced apoptosis. Namely, the E1B and E4orf6 helper proteins (encoded by *E1b* and *E4orf6*, respectively) are able to destabilise p53 by forming a complex with associated cellular proteins. This complex harbours an E3 ubiquitin ligase function and targets p53 for ubiquitin-mediated proteolysis (Luo *et al.*, 2007; Schwartz *et al.*, 2008). Furthermore, E1B and E4orf6 possess additional helper functions that promote the nuclear export of late viral mRNA to the cytoplasm for downstream processing (Pilder *et al.*, 1986; Krätzer *et al.*, 2000; Blanchette *et al.*, 2008). Considering AAV or Adenovirus virus-specific elements are exposed within infected cells, the host cellular anti-viral responses must be repressed. The VA RNAs assist in this regard to impressively block anti-viral mechanisms by

inhibiting interferon (IFN) signalling that would otherwise activate anti-viral enzyme systems such as Dicer and the RNA-induced silencing complex (RISC) (Andersson *et al.*, 2005). Lastly, the final helper function required for AAV production is performed by E2A (encoded by *E2a*), which functions to promote AAV DNA replication (Ward *et al.*, 1998). Therefore, the range of helper functions work in concert to promote an intracellular milieu permissive for AAV2 production.

1.2.7 AAV site-specific integration

On the other hand, in the absence of a helper virus or helper functions, AAV is significantly compromised in self-sufficiency and undergoes latency. Here, AAV genome replication is limited and gene expression is repressed. Latency is established by the preferred site-specific integration of AAV genome into chromosome 19q13.4, dubbed the AAV integration site 1 (*AAVS1*) (Kotin *et al.*, 1990; Samulski *et al.*, 1991). Specifically AAV genome integrates into the initial exon of the myosin binding subunit 85 (*MBS85*) gene (Janovitz *et al.*, 2013). This site-specific integration accounts for roughly 45% of all integrations detected by Janovitz *et al.* (2013).

A direct consequence on the lack of helper virus co-infection is the limited expression of AAV Rep78 and Rep68. This results in the repression of AAV-specific gene expression and DNA cannot be replicated (Labow and Berns, 1988). Integration into the *AAVS1* locus is particularly favoured because of the presence of GCTC repeat elements, of which elements appear similarly in wildtype AAV *ITRs*. The presence of these elements permits the N-terminus of Rep78 and Rep68 proteins to bind to (Weitzman *et al.*, 1994; Surosky *et al.*, 1997). Additionally, the presence of a *trs* within the *AAVS1* site also has been identified, and thought to be nicked upon integration of AAV genome (Linden *et al.*, 1996). In particular, supplying the *rep* gene in *cis* to an *ITR*-flanked GFP expression cassette demonstrated enhanced integration into *AAVS1* compared to supplying in *trans* (Balagúe *et al.*, 1997). Given that the only viral components present in these investigations included the *rep* gene and *ITR* sequences, this essential work showed that the minimum requirements for site-specific integration was the presence of Rep and *ITRs*.

The exact mechanism of AAV integration has been explored, especially to elucidate any adverse consequence of AAV integration or latency. Rep78/68 protein are capable of binding to both AAV RBE and cellular RBE at the same time (Weitzman *et al.*, 1994), suggesting a

possible mechanism of integration involving Rep tethering the *ITR* containing sequence to the *AAVS1* site. Henckaerts *et al.* (2009) investigated site-specific integration using mouse embryonic stem cell lines because mice cells harbour an *MBS85* orthologue. It was observed that integration of AAV genome was associated with duplication of the target integration site. Specifically, and given integration can occur with the AAV genome orientated in the 5'-3' transcriptional direction of the *Mbs85* gene (Lusby *et al.*, 1981; Rahim *et al.*, 2011), the left *ITR* sequence was found within the *Mbs85* gene. Whereas the right *ITR* was located in the promoter region of *Mbs85* instead, with the addition of 18nt spare that resembled the reverse complement of the *Mbs85* sequence from the junction that reads into the left *ITR* (Henckaerts *et al.*, 2009). Therefore, integration was suggested to be mediated by Rep, which nicks at the cellular *trs* in *AAVS1* leading to DNA synthesis of the target site with co-recruitment of AAV genome via Rep and AAV left *ITR* RBE sequence. With AAV genome in proximity, strand displacement leads to AAV genome being replicated instead and is contiguous with the initial replication of the target integration site. Seeing as 18nt (reverse complement) of the *Mbs85* was observed by Henckaerts *et al.* (2009) at the far right junction at the integration site, this suggested that the AAV genome was replicated and the replication fork switches back to the *Mbs85* sequence (the strand complementary to the displaced strand) as template. The proposed model (Fig. 1.4) is extended by ligation of the 3'-hydroxyl group of the nascent strand with the 5'-end of the originally displaced cellular DNA by Rep, because Rep has shown previous ligase activity (Smith and Kotin, 2000). Although, the importance of cellular ligase IV in AAV integration was more recently noted by Daya *et al.* (2009), where significantly less AAV specific integration was observed in cells that expressed less ligase IV. Nonetheless, quiescent AAV is readily rescued when latently infected cells are subsequently infected by helper virus. The lytic life-cycle of AAV commences with the excision of AAV provirus from the host genome (a Rep-dependent process), and the production of infectious progeny in the presence of superinfection with Adenovirus (Berns *et al.*, 1975).

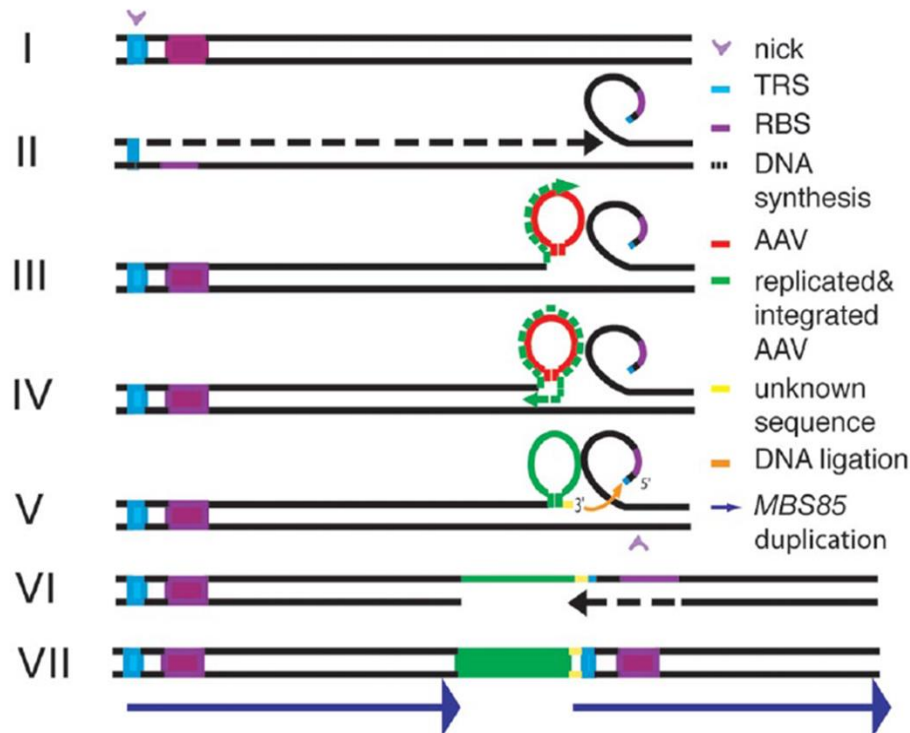


Figure 1.4 Model for AAV site-specific integration into AAVS1 locus. I) A nick is generated in the trs structure in AAVS1, mediated by AAV Rep78/68. II) Strand displacement occurs by DNA synthesis at site of nick. III) Strand displacement occurs with AAV genome as template instead. IV) DNA synthesis switches back to the *Mbs85* DNA sequence (complementary to the displaced strand). V) Ligation between synthesised DNA (here annotated as replicated AAV and ‘unknown sequence’) and the displaced DNA strand occurs. VI) A nick is generated in the opposite strand by Rep, and DNA synthesis completes AAV genome integration. VII) This model of AAV integration into the AAVS1 generates a duplication of the *Mbs85* sequence. Figure from Henckaerts *et al.* (2009).

1.2.8 AAV receptor-ligand interaction

Infection by AAV is a multi-step process that involves virus attachment to relevant cell surface receptors. This leads to intracellular signalling that promotes virus uptake by endocytosis. AAV is then trafficked towards the nucleus, and its ssDNA genome is translocated into the nucleus for replication and gene expression for production of progeny. AAV2 has been shown, by single virus tracing studies by Seisenberger *et al.* (2001), to make multiple contacts with cell membranes to decelerate. By virtue of AAV2’s selective binding to its primary attachment receptor, heparan sulphate proteoglycan (HSPG) (Summerford and Samulski, 1998; Rabinowitz *et al.*, 1999), AAV2 is capable of making the necessary contacts it requires to slow down its acceleration and firmly attach to cell surface membranes. This process seems especially reliant on Vp3 capsid protein (Rabinowitz *et al.*, 1999). The use of site-directed mutagenesis of the cap ORF helped identify critical functional domains within Vp3, which facilitate receptor-ligand binding between AAV2 and HSPG (Wu *et al.*, 2000). The corresponding mutations introduced by Wu *et al.* (2000) correlated with two amino acid (aa) clusters in Vp3’s loop IV. The precise

aa residues that contribute to the heparin-binding motif was further elucidated by Kern *et al.* (2003) as being R484, R487, K532, R585, and R588. These are located as patches of basic aa within loop IV at the three-fold rotation axis of AAV2 capsid. Considering that these key residues are not exposed on the capsid surface but rather buried, and that K532A and R585M mutants showed only a modest effect in AAV2 infectivity, it seems that AAV2 binding to HSPG initiates contact to cell membranes in order to enhance cell-virus interactions for anchorage, and is not wholly necessarily for infection.

It was also apparent that a number of co-receptors were necessary to facilitate AAV2 entry. It was identified that co-receptors $\alpha V\beta 5$ integrin, $\alpha V\beta 1$ integrin (Asokan *et al.*, 2006), fibroblast growth factor receptor-1 (FGFR1) (Qing *et al.*, 1999), and hepatocyte growth factor receptor, c-Met (Kashiwakura *et al.*, 2005) enhanced infectivity with little to no effect on the binding properties of the virus particles to target cell. For instance, cells treated with antibodies that targeted $\alpha V\beta 5$ integrin showed inhibited endocytosis of AAV2, but binding affinity was unchanged relative to the use of control anti-mouse immunoglobulin G (IgG) (Sanlioglu *et al.*, 2000). However, it was noted that initial attachment of virus to cell surface seemed to be enhanced by the FGFR1, as the loss of FGFR1 expression alone correlated with the inability for AAV2 to bind (Qing *et al.*, 1999).

1.2.9 AAV cell entry

As a consequence of AAV binding to primary cell surface receptor (HSPG) and secondary co-receptor(s), downstream signalling pathways are activated to permit virus uptake by clathrin-mediated endocytosis (Bartlett *et al.*, 2000). The sequestration of virus by αV integrins and localisation of AAV2 to clathrin-coated pits facilitates the molecular cue for clathrin-mediated endocytosis (Mukherjee *et al.*, 1997; Wang *et al.*, 1998). This paradigm of cell entry is conserved between AAV serotypes (Bartlett *et al.*, 2000). Even serotypes that bind to an alternative, unrelated primary receptor(s), such as AAV5 (which binds to sialic acid receptor), has also been shown to sequester to clathrin-coated pits for receptor-mediated endocytosis (Bantel-Schaal *et al.*, 2002).

Downstream signalling events associated with actin cytoskeletal reorganisation and endocytosis has been largely supported. AAV2 has been shown to internalise into HeLa cells in a Ras-related C3 botulinum toxin substrate 1 (Rac1)-dependent manner (Sanlioglu *et al.*, 2000). Rac1 is a small signalling GTPase involved in cell cytoskeletal reorganisation, cell

motility, adhesion and membrane trafficking (Lamaze *et al.*, 1996; Ridley, 2006; Fujii *et al.*, 2013). Integrins have been widely known to interact with a number of intracellular signalling molecules and function as molecular cues for downstream signalling events. Integrin interactions with Rac, Rho and Cdc42 families of GTPases have been described (Nobes and Hall, 1995; Van Aelst and D'Souza-Schorey, 1997). With the clustering of integrins to a focal point at the cell surface enables proximal activation of phosphatidylinositol-3 kinase (PI3K) pathway, followed by Rac1 activation (Li *et al.*, 1998). PI3K is a lipid kinase which phosphorylates PtdIns(4,5)P₂ (PIP₂) to form the active secondary messenger, PIP₃. PIP₃ then acts upon small GTP-binding proteins such as Rac1 and results in actin polymerisation. Ultimately, actin filaments propagate directly under the plasma membrane at the focal point. Membrane extensions then form and engulf AAV2 particles in the initial process of endocytosis. Complete internalisation is deemed dynamin-dependent, where dynamin oligomerisation forms a ring structure that is necessary for the formation of clathrin-coated vesicles, and the final pinching of coated pits from the cell membrane into the internal compartment (Duan *et al.*, 1999). However, inhibition of dynamin showed only a partial block in viral endocytosis (Duan *et al.*, 1999), indicating the presence of alternative entry mechanisms. It should be noted that a number of the aforementioned investigations assayed AAV2 infection in the presence of Adenovirus co-infection for helper function, which may affect cell homeostasis and thus confound previous reports on AAV2 internalisation.

On the other hand, Nonnenmacher and Weber (2011) demonstrated that infectious AAV2 internalisation was predominantly controlled by the clathrin-independent carriers (CLIC)/GPI-anchored-protein-enriched endosomal compartment (GEEC) pathway, instead. This was determined by blocking clathrin-mediated endocytosis through the overexpression of truncation mutant of Eps15 (a clathrin-coated vesicle component), or chemical inhibition using chlorpromazine. Virus internalisation was unaffected under these conditions indicating AAV2's independence to clathrin-mediated endocytosis. Alternatively, inhibiting effectors of CLIC/GEEC, such as Arf1 and Cdc42 GTPase showed reduced transduction of AAV2 into HeLa and 293T cells by 70%. AAV2 was evidenced within CLIC/GEEC endosomes, as per co-localised markers of CLIC/GEEC endosomes, such as GRAF1 (Nonnenmacher and Weber, 2011). AAV2 entry was not completely blocked by inhibiting clathrin-mediated endocytosis mentioned above, whereas, inhibition of both the CLIC/GEEC- and dynamin-dependent pathways of endocytosis were needed to block entry completely. Therefore, both entry pathways are simultaneously used by AAV2 (Fig. 1.5), although the dependence of either may vary

considerably between different host cell types.

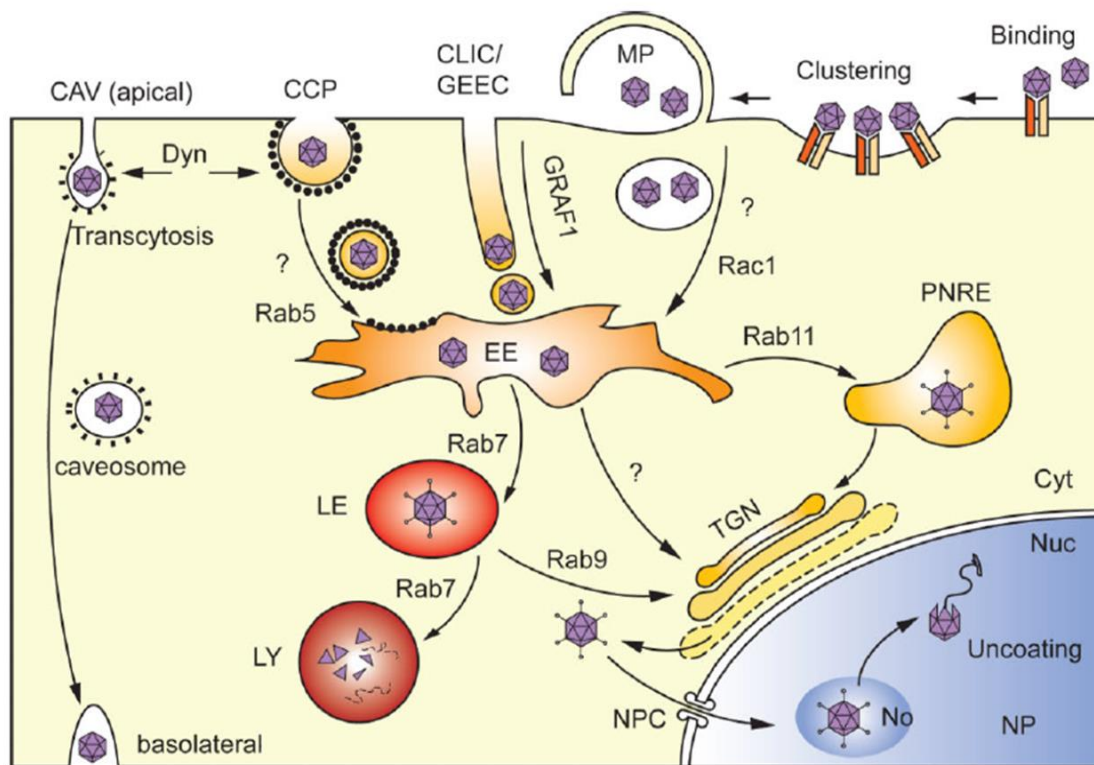


Figure 1.5 Proposed mechanisms of AAV vector trafficking to the nucleus. AAV binds to key receptors on the surface of target cells, and are internalised by clathrin-mediated endocytosis (CCP) and/or CLIC/GEEC. Trafficking of AAV then is transported to the nucleus, likely involving early endosome maturation to late endosome or recycling endosomes. AAV is then released to transverse the nuclear membrane into the nucleus, to continue its life-cycle. Acronyms thereafter include: CAV, caveolar endocytosis; Cyt, cytosol; Dyn, dynamin; ER, endoplasmic reticulum; EE, early endosome; LE, late endosome; LY, lysosome; MP, micropinocytosis; No, nucleolus; NP, nucleoplasm; NPC, nuclear pore complex; Nuc, nucleus; PNRE, perinuclear recycling endosomes; TGN, trans-Golgi network. Figure from Nonnenmacher and Weber (2012).

1.2.10 AAV trafficking to the nucleus

Inarguably, the AAV genome must be trafficked into the nucleus, express relevant viral proteins, and replicate progeny in order to establish an infective or latent life-cycle. Successful transduction of AAV requires intact particle to migrate to the nucleus, and import at least its genome. It remains unclear exactly how AAV is trafficked to the nucleus, especially when exactly viral uncoating occurs. To answer these, studies have predominantly utilised rAAV vectors and examined the subcellular localisation and trafficking mechanisms employed.

Recombinant AAV vectors are compartmentalised into Rab5⁺ early endosomes after initial uptake by clathrin-mediated endocytosis. These vesicle-bound rAAV are shown to traffic to the perinuclear space (Bartlett *et al.*, 2000; Sanlioglu *et al.*, 2000). More recent evidence suggests AAV is trafficked to the Golgi apparatus that occupies the perinuclear space (Bantel-Schaal *et*

al., 2002; Xiao and Samulski, 2012; Xiao *et al.*, 2016). Strong evidence suggests that compartmentalised AAV utilises the cells' microtubule network to traffic itself towards the nucleus via the perinuclear space, unidirectionally (Xiao and Samulski, 2012). Despite using nocodazole treatment to disrupt microtubule formation and inhibit the formation of the microtubule organisation centre (MTOC), a small portion of infecting AAV still manage to translocate into the nucleus, indicating alternative mechanisms for AAV trafficking and/or nuclear translocation (Xiao and Samulski, 2012).

Efficient trafficking and transduction is dependent on early endosome processing to late stage endosomes. Reports have shown trafficking of AAV through the Rab7⁺ late endosomes; although trafficking via Rab11A⁺ recycling endosomes has also been reported (Douar *et al.*, 2001; Ding *et al.*, 2006; Harbison *et al.*, 2009). Redirecting AAV to traffic via the Rab7⁺ vesicles, using Eer1 to inhibit endosomal reticulum-associated degradation processing, contributed to enhanced transduction efficiencies (Berry and Asokan, 2016). This indicated a benefit for AAV particles to commit to a single trafficking pathway. Interestingly, a dose-dependent disparity in AAV2's infection route has been described: high multiplicity of infection (MOI) of 10,000 genomes/cell correlated with co-localisation of AAV2 particles with Rab11A⁺ endosomes. Rab7A⁺ late stage endosomes was associated with low MOI of 100 genomes/cell (Ding *et al.*, 2006). In addition to this, the mechanism utilised by AAV to traverse infected cells likely is cell-type dependent (Pajusola *et al.*, 2002), as chemical inhibition of micropinocytosis in HeLa, HepG2 or Huh7 cell lines showed enhanced transduction rates only in the hepatocellular cell lines (Weinberg *et al.*, 2014). Differential trafficking pathways may be a feature of different AAV serotypes infecting the same permissive cells. This may be attributed to the fact that different AAV serotypes utilise different primary receptors to mediate viral or vector entry into target cells, which would likely impact on downstream signalling events (Liu *et al.*, 2013).

Endosomal escape of rAAV typically ensues, and involves the release of rAAV from their processed endosome vesicles. Therefore, once trafficked close to the nucleus, endosomal pH is modulated to favour AAV escape. Endosome maturation into late endosomes and lysosomes is accompanied with acidification of this compartment. The use of chemical inhibition of endosomal acidification has outlined the importance of this step. Inhibitors have included ammonium chloride, bafilomycin A1, or chloroquine, all of which treatments correlated with reduced AAV transduction (Bartlett *et al.*, 2000; Xiao and Samulski, 2012). Eventually, a conformational change in the N-terminus of Vp1 of AAV capsid, which houses the

phospholipase A₂ (PLA₂) domain and nuclear localisation signals (Fig. 1.6), is stimulated (Venkatakrishnan *et al.*, 2013). Given both the inhibition of late-endosome formation and acidification of late endosomes is associated with reduced AAV transduction, it seems late-endosome acidification is the trigger for this conformation change. Mutations in PLA₂ catalytic domain correlated with reduced AAV gene expression and translocation into the nucleus, but did not affect AAV entry into cells and its localisation to the perinuclear space (Girod *et al.*, 2002; Stahnke *et al.*, 2011). Moreover, AAV infectivity and transduction were rescued when co-infecting Vp1 mutated AAV vectors with specialised AAV vectors, which were composed of Vp1 fusion proteins that allowed their capsid surfaces to be exposed, with inherently activated PLA₂ domains (Grieger *et al.*, 2007). In turn, these data imply that the PLA₂ domain is necessary for AAV infectivity downstream of AAV entry and trafficking to the perinuclear space. Given PLA₂ functions to hydrolyse phospholipids, it was reasonably assumed that this activity is employed by infectious AAV to hydrolyse the endosome compartment's membrane and escape.

It remains widely accepted that AAV translocates into the nucleus as intact particles (Nicolson and Samulski, 2014; Kelich *et al.*, 2015). Despite the growing evidence in support of nuclear translocation of intact AAV, the point at which viral uncoating actually occurs remains debated. Fewer studies imply viral uncoating of AAV2 as occurring prior to nuclear translocation (Lux *et al.*, 2005). In further support of the former is the presence of three basic regions (BR1-3) in the unique N-termini of Vp1 and Vp2 (Fig. 1.6). The sequences of which closely resemble nuclear localisation signals (NLS) (Grieger *et al.*, 2006). Mutations in BR1-3 are associated with reduced transduction and the inability for AAV to translocate into the nucleus (Johnson *et al.*, 2010). The mechanism in which intact AAV particles physically translocate the nuclear membrane has thus been proposed, and seems to operate in a different manner to minute virus of mice (MVM) – another parvovirus – that disrupts the nuclear lamina in order to access the nucleus (Cohen *et al.*, 2006; Cohen *et al.*, 2011). The proposed mechanism has been shown to involve nuclear pore complexes (NPCs) instead for AAV2. Labelled AAVs were directly evidenced to move into the nucleus through labelled NPCs (Kelich *et al.*, 2015). And the inhibition of NPCs using wheat germ agglutinin prevented nuclear entry of infecting Cy5-AAV to a greater extent than vehicle control used in the study (Nicolson and Samulski, 2014). Furthermore, rAAV2 was found to co-localise with importin- β (a key chaperone involved in directing nuclear entry of NLS containing proteins) after 1h post-infection, and post-acidification of AAV2 containing endosomes (Nicolson and Samulski, 2014).

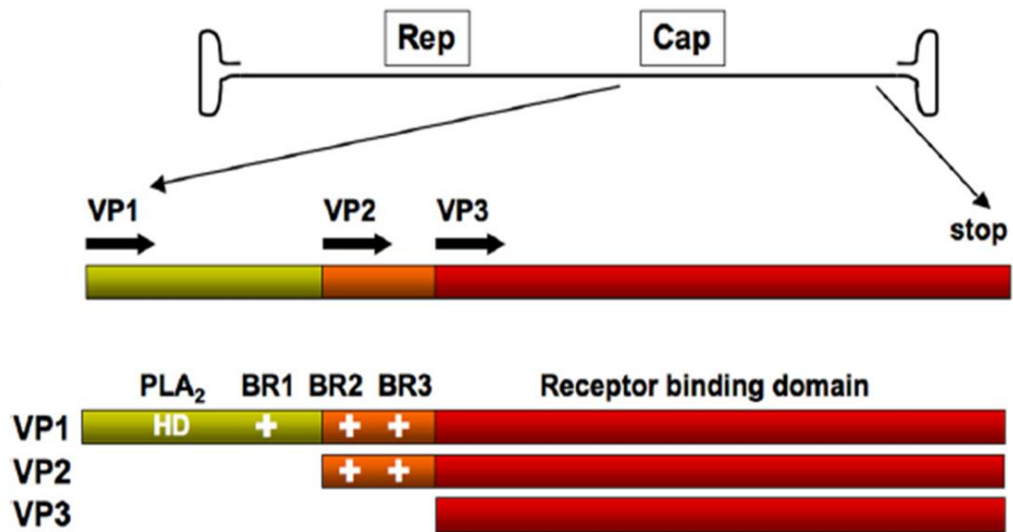


Figure 1.6 Vp1 and Vp2 contain key signals for AAV life-cycle. Schematic of AAV Vp1-3 proteins, which contain PLA₂ (HD) domain and basic regions (BR1-3) in the N-terminus of Vp1 and Vp2 proteins necessary for AAV infectivity. + regions indicate putative nuclear localisation signals. Figure from Johnson *et al.* (2010).

After entering the nucleus AAV2 localises to nucleoli, but mobilisation into the nucleoplasm seems to be required for AAV to uncoat and permit gene expression (Johnson and Samulski, 2009). The exact means that intact AAV capsids use to traffic to nucleoli is not recognised in full, but nucleolar proteins have been shown to directly interact with intact AAV. Namely, nucleolin and nucleophosmin have been shown to bind and co-localise with AAV after nuclear import (Qiu and Brown, 1999; Bevington *et al.*, 2007). Although, nucleolar accumulation of infecting AAV seems to act as a barrier to efficient transduction; specifically short interfering (si)RNA knockdown of either nucleolin or nucleophosmin in HeLa cells showed enhanced transduction efficiencies (Johnson and Samulski, 2009). Chemical treatment with hydroxyurea significantly affected the subcellular distribution of nucleolin, and treatment with hydroxyurea correlated with enhanced AAV transduction efficiency, with a diffuse distribution of AAV in the nucleoplasm but excluded from the nucleoli (Johnson and Samulski, 2009).

1.2.11 AAV DNA replication

AAV genome replication predominantly occurs in the presence of Adenovirus or HSV co- or super- infection, which provides all the necessary helper functions that permit AAV transcription and gene expression. Ultimately, AAV DNA replication occurs by the synthesis of dsDNA from the ssDNA AAV genome acting as template. The dsDNA AAV replication product serves as replicative intermediates (Straus *et al.*, 1976), from which ssAAV genomes with plus

or minus polarities are packaged into AAV capsids (Berns and Rose, 1970). Host cell machineries and viral factors require access to naked AAV genome in order to commence DNA replication. Although, the disassembly of AAV capsid has not been fully elucidated, work by Johnson and Samulski (2009) propose infectious AAV2 vector remain compartmentalised in nucleoli, and mobilise to the nucleoplasm in order to instigate AAV genome replication or gene expression.

DNA replication is achieved initially from the inherent property of the AAV *ITRs*. The *ITRs* harbour the viral origin of replication in the forms of the RBE and *trs*, but also, the initial-most 125nt of *ITRs* form imperfect palindromes that folds upon itself. This self-annealing property provides a suitable dsDNA template, and importantly provides a base-paired 3'-hydroxyl group from which unilateral DNA synthesis can occur. The replication machinery is thought to be recruited for unidirectional DNA synthesis from the *ITR*-derived 3'-hydroxyl group. However, the unilateral DNA replication alone does not complete DNA replication of the entire AAV genome, due to a lack of a base-paired 3'-hydroxyl group on the opposite strand to replicate the remainder of 3' *ITR* that served as a replication primer. In this case, AAV Rep68/78 proteins bind to the dsDNA of the *ITR* via the RBE (McCarty *et al.*, 1994). Rep also recognises a second motif (RBE') present in the top of one of the *ITR*'s hairpins (Hickman *et al.*, 2004). Rep is thought to function as a helicase to generate ssDNA around the *trs*, which is necessary to allow the *trs* to form an intermediary hairpin structure. This change in conformation within the AAV *ITR* sequence promotes cleavage of the *trs* via Rep's endonuclease function, and provides another 3'-hydroxyl group to permit DNA synthesis and single-strand displacement. The remainder of the *ITR* that served as a replication primer is faithfully replicated, resulting in a completely replicated AAV genome as a dsDNA intermediate. The DNA ends of which can renature to form terminal hairpins from each strand and generate another base-paired, 3'-hydroxyl group for replication by single-strand displacement. This provides concatemeric AAV genomes and the ssAAV genome for packaging (Straus *et al.*, 1976; Hong *et al.*, 1994). An overall schematic of replication by single-strand displacement is depicted below in Fig. 1.7.

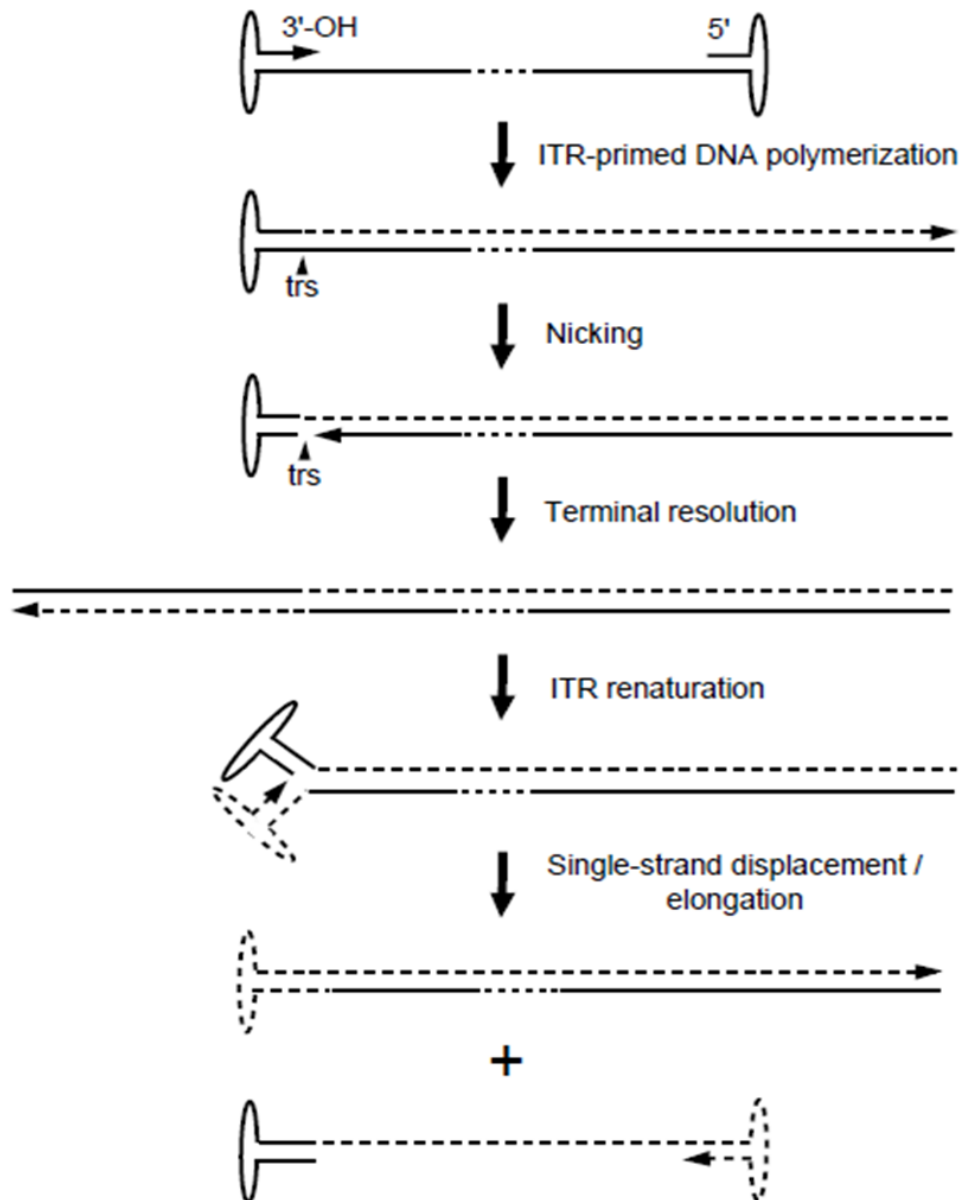


Figure 1.7 Model for AAV genome replication. The 3' *ITR* of AAV2 genome forms a T-shaped hairpin structure, which provides a base-paired, 3'-hydroxyl group necessary for DNA synthesis. Firstly Rep68/78 binds to the 3' *ITR* as the origin of replication, and further recruitment of the cell's replication machinery facilitates DNA synthesis of the AAV genome. The *trs* site at the 3' of the AAV genome is nicked by Rep68/78, and provides another base-paired, 3'-hydroxyl group to permit complete DNA replication of the *ITR* that served as a replication primer. After complete DNA replication, the *ITRs* can renature to serve as replication primers to promote single-strand displacement and elongation of the AAV genome. The entire model results in concatameric dsAAV genome. The ssAAV genomes can be packaged. Solid lines refer to template strand; dashed lines refer to replicating strands. Figure from Goncalves (2005).

1.2.12 AAV assembly

It is commonly perceived that AAV assembly occurs in two distinct phases, i) assembly of a preformed capsid shell, followed by ii) the packaging of ssAAV genome into the preformed capsid (Myers and Carter, 1980). The assembly of preformed capsid vectors has been observed

in the nucleoli or nucleolar periphery, with encapsidation of AAV genome occurring in the nucleoplasm (Wistuba *et al.*, 1997). The expression of Vp1-3 leads to the coordinated assembly of AAV capsid with the 1:1:10 stoichiometry of Vp1, Vp2 and Vp3, respectively. This process of which requires host factors in order to facilitate capsid assembly (Steinbach *et al.*, 1997; Wistuba *et al.*, 1997).

Assembly of preformed AAV capsids is promoted by an additional protein, AAP, which is encoded by an alternative ORF in the *cap* gene (Fig. 1.1). Expression of Vp2/3 and Vp3 only capsid proteins can result in successful assembly of empty vectors, albeit, non-infectious vectors (Grieger *et al.*, 2007). This is likely due to the loss of the PLA₂ domain and BR1-3 sequences. Nonetheless, vectors derived from solely Vp3 expression was only possible when the entire AAV2 *cap* gene, with mutated initiation codons for Vp1 and Vp2 ORFs, was transfected (Sonntag *et al.*, 2010). This indicated the presence of an AAV-encoded factor upstream of Vp3 ORF responsible for nucleolar trafficking and capsid assembly of Vp3, and was characterised as AAP (Sonntag *et al.*, 2010; Sonntag *et al.*, 2011; Earley *et al.*, 2015). Silencing AAP expression prevented the nucleolar trafficking of Vp1-3 proteins and capsid formation, which could not be rescued by tagging Vp3 with a nucleolar signalling peptide (Sonntag *et al.*, 2010). Therefore, a proposed role for AAP is to function as a scaffolding protein by binding to capsid proteins and concentrating the assembly process to the nucleoli (Naumer *et al.*, 2012; Earley *et al.*, 2015). The mechanistic role of AAP in promoting capsid assembly may be conserved between serotypes given AAV2-derived AAP can promote capsid assembly of a diverse range of AAV serotypes (Sonntag *et al.*, 2011; Naumer *et al.*, 2012). However, Earley *et al.* (2015) have shown that AAP is not essential for capsid assembly for AAV4, 5, and 11, and for nucleolar localisation of capsid assembly.

Encapsidation of ssAAV genome into preformed capsids is likely to follow after AAV genome is replicated, and is thought to occur in the nucleoplasm at later stages of AAV infection (Wistuba *et al.*, 1997). Additionally, Rep -40, -52, -68, and -72 proteins were shown to bind directly to free capsid proteins or assembled capsids independent of *ITR*-containing AAV vector DNA, by co-immunoprecipitation studies (Dubielzig *et al.*, 1999). Certain bulky mutations at AAV2's pores at their five-fold symmetry axes correlated with reduced Rep-capsid interactions and genome packaging efficiency (Bleker *et al.*, 2005; Bleker *et al.*, 2006). Such protein-protein interaction may generate an intermediate complex with AAV ssDNA, given Rep78 covalently associates with AAV 5'-*ITR* sequence (Prasad and Trempe, 1995).

AAV vector genomes have been identified to enter AAV capsids in a polar manner. The 3'-ssDNA sequence was identified in partially-packaged AAV2 vectors implying the 3'-ssDNA sequence is first translocated into preformed capsids (King *et al.*, 2001), despite the fact that the 5'-sequence associates with preformed capsids in context with Rep78 (Prasad and Trempe, 1995). Especially, the loss of the *ITR*'s D-sequences, but retention of the hairpin secondary structures, in plasmid constructs was defective in their ability to produce ssDNA genomes. Therefore, ssDNA progeny genome could not be packaged, and vector genomes were not detectable (Wang *et al.*, 1996). Further to this, after associating dsAAV genomes to preformed capsid via Rep78, smaller Rep40/52's DNA helicase activity promoted the unwinding the ds AAV genomes, principally at the 3'-end to feed and encapsidate ssAAV genome (King *et al.*, 2001).

1.3.0 *In vitro* production of recombinant AAV vectors

1.3.1 Recombinant AAV vectors

The development of rAAVs for vector-based gene therapy has progressed significantly and continuously over the years. Use of AAV vectors has become a very successful and promising therapeutic tool for treatment of monogenetic diseases. However, despite the current developments, certain limitations have restricted the full utilisation of AAVs. One main limitation is the constrained viral titres produced from current production methods. Interest in this field of research is due to the need for high MOI for efficient *in vivo* transduction by rAAV vectors. Generally an MOI of 10^3 - 10^5 infectious particles/cell is predicted to be needed (Ellis *et al.*, 2013), or clinical doses of 10^{12} - 10^{13} /kg infectious particles needed for human administration (Bryant *et al.*, 2013); though these quantities are quite dependent on the intended target tissue type.

1.3.2 Components required for rAAV production

Recombinant AAV production is mediated most commonly by triple transfection of three recombinant plasmids that introduce the transgene, *cis* and *trans* components of AAVs, as well as helper functions. These recombinant plasmids individually encode for the transgene of interest flanked by AAV2 *ITRs*, or the Rep and Cap proteins, or finally the helper functions minimally required for rAAV vector production (Matsushita *et al.*, 1998; Xiao *et al.*, 1998). The gold standard for the laboratory-scale *in vitro* production of vectors is the use of the packaging cell line, human embryonic kidney (HEK)293T. Therefore, adherent 293T cells are subjected to triple transfection by above mentioned plasmids to commence rAAV production.

293T cells are a specialised cell line, and have been developed and commercialised for the use in rAAV vector production. This immortalised cell line has been specifically modified by the addition of human Adenovirus type 5 *E1* early region coding DNA, and in turn provides E1A and E1B helper function (Graham *et al.*, 1977). Therefore, an advantage of this transformation allows for a immortalised cell line as per E1A oncogene function, and the necessary helper function provided by E1A's *trans*-acting function to up-regulate viral promoters. Another advantage of using a 293-AAV specialised cell line is the partial independence from Adenovirus or alternative helper virus infection, which improves the safety profile of subsequent rAAV vectors – the risk of pathogenic contamination is ablated. The remainder minimal helper function is provided by the aforementioned helper plasmid that provides the E2A, E4orf6 proteins and the VA RNA's. This is delivered into the packaging cell line simultaneously by triple transfection with the AAV-transgene and *rep/cap* plasmids.

It should be noted that the transgene of interest is cloned into a vector backbone that comprises the AAV2 *ITRs*, and therefore the *rep* and *cap* ORFs have been replaced completely. With the removal of the *rep* and *cap* coding sequences with intended transgene, the coding capacity increases significantly. A coding capacity of approximately 5kb in length is generated with the replacement of the *rep* and *cap* genes. The *ITRs* are maintained for their *cis*-acting functions in particle assembly (packaging of the recombinant AAV genome into preformed empty capsids), and genome replication. Additionally, the discovery of different serotypes of AAV (Rutledge *et al.*, 1998) promoted pseudotyping of rAAV vectors. This meant the capsid encoding element of the recombinant systems derived from different serotypes, whereas the *ITRs* and *rep* gene were based on AAV2. This permitted a wide range of serotypically different vectors to be produced, and thus expanded the transduction potential of rAAV vectors because pseudotyping particles provide a wide collective tropism.

1.3.3 Recombinant AAV vector production

Scalable systems of rAAV production are currently in development. However, an important consideration to take into account is the need to produce sufficient quantities of vector that can be used for clinical use. This has been a long-standing limitation in the history of rAAV-mediated gene therapy, but constant advancements have been developed to procure higher and purer yields of rAAV.

Generally speaking 293T cells are grown and expanded *in vitro*. After which, triple transfection

with three plasmids (encoding transgene and *cis*- or *trans*- acting factors) is performed on adherent 293T cells (Fig. 1.8). A number of means can be utilised to promote the efficient transfer of plasmid DNAs into 293T cells, especially the use of Calcium Phosphate precipitation (Xiao *et al.*, 1998), and polyethylenimine (PEI) (Drittanti *et al.*, 2001), and cationic lipids (Liu *et al.*, 2008). The Calcium Phosphate precipitation method is widely used for large-scale production of rAAV, attributable by the reagents cost-effective use in transfection of 293T cells, especially with high transfection efficiency documented by Meissner *et al.* (2001) (up to 90% transfer of exogenous DNA into cultured cells). It must be noted that for rAAV vector production using Calcium Phosphate or even other chemical methods of DNA transfer, the process and conditions must be specially optimised. Additionally, consistency between productions must be upheld to truly ensure reproducible and comparable AAV titres and quality. Additional considerations are the viability and quality of cultured 293T cells and the variation in protocols between established laboratories. Consistency in this respect is difficult to achieve, with variation in chemical makeup of reagents and technique, culture conditions, all of which can impact on rAAV vector production. Despite this, Calcium Phosphate-mediated triple transfection has shown to provide vector titres ranging from 10^3 - 10^5 viral genomes/cell (Aucoin *et al.*, 2008), therefore showing a range of 100-fold difference in infectious viral genomes. This is a considerable difference to have to account for, and one that has proven difficult to narrow.

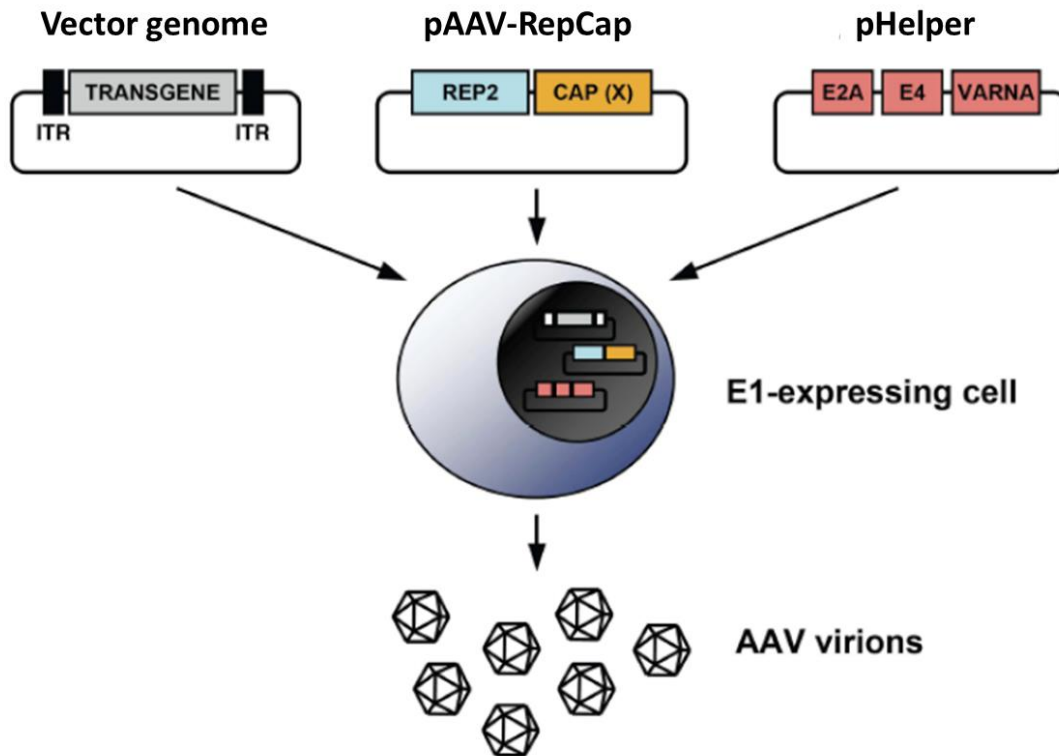


Figure 1.8 Schematic of a common AAV vector production method by triple transfection. The triple plasmid system to generate rAAV vectors in 293T, or suitable E1A/B-expressing cell line is illustrated. The vector genome is encoded by one plasmid, and contains the gene of interest under transcriptional control of a suitable promoter and polyA signal, all of which is flanked by AAV2 *ITRs*. The second plasmid (pAAV-RepCap) encodes AAV2 Rep protein, and Cap proteins for the desired AAV serotype. The third plasmid (pHelper) contains E2, E4 and VA-RNA – the minimal adenovirus factors required for AAV replication. 293T cells are ultimately co-transfected, after which, AAV vectors can be harvested. Figure adapted from Ayuso *et al.* (2010).

Nonetheless, triple transfection results in the production of rAAV and 72h post-transfection, cells are harvested and lysed to recover rAAV vectors. Therefore, each cell treated must be transfected with each of the three plasmids for vector production to occur. This transfection pressure therefore restricts the efficiency of production on a cell-by-cell basis. Subsequent purification of infectious rAAV vectors from empty capsids and cellular proteins is then performed. For example purification was most initially based on multiple rounds of caesium chloride density-gradient ultracentrifugation. Here, caesium chloride salts are able to form gradients based on differential buoyancies. This generates differing equilibriums between sample compositions when subjected to ultracentrifugation. When viruses are treated with caesium chloride salts and centrifuged, viruses are collectively separated from contaminants and cellular constituents. In fact, where rAAV production involves Adenovirus helper, caesium chloride density-gradient ultracentrifugation is capable of partitioning rAAV from Adenovirus contaminants. Main disadvantages of this purification method must be considered; these include caesium chloride toxicity, which means downstream clinical use is limited unless

viruses are dialysed against a physiological buffer. Further to this, reports have indicated that rAAV stocks purified by this method have shown loss in purity and infectivity (Zeltner *et al.*, 2010; Strobel *et al.*, 2015).

An alternative purification method involves non-ionic iodixanol gradients, and was developed by Zolotukhin *et al.* (1999). The bulk purification step by iodixanol provides a number of benefits, including faster turnaround for purifying rAAV, better infectivity of rAAV particles, and lower downstream toxicity compared to caesium chloride (Zolotukhin *et al.*, 1999; Potter *et al.*, 2002; Zolotukhin *et al.*, 2002). Additionally, aggregation of rAAV particles was not a feature associated with iodixanol density-gradient purification, and the non-ionic and non-viscous characteristic of the iodixanol medium meant that preps could be loaded onto chromatographic columns almost directly. Chromatographic steps also resulted in higher yields of infectious titres when crude lysates were subjected to iodixanol density gradient purification, compared to using caesium chloride gradients by up to 10-fold. An impressive 70% of rAAV from crude lysates could be recovered (Zolotukhin *et al.*, 1999). Following this initial step of purification, rAAVs are subjected to further purification by affinity chromatography. This technology exploits AAVs natural ligand for receptor binding. HSPG has been shown to facilitate AAV2 entry upon receptor binding (Summerford and Samulski, 1998), and shows high sensitivity and affinity of binding. Therefore use of high-performance liquid chromatography (HPLC) with heparin columns can purify rAAV vector products. A purity of >99% was documented by Zolotukhin *et al.* (1999). This method of purification is also limited by the selectivity of binding – only rAAVs with a tropism for heparin will bind in the above setup. Alternative to this, the use of capsid specific antibodies conjugated to heparin columns bypasses the purification restriction. In this way a broader class of AAV serotypes can be purified for a wider set of downstream uses.

1.3.4 Advancements in rAAV vector production methods

The traditional production systems involved a double-transfection setup (using plasmids that encoded for AAV2 *ITRs* and transgene, or *rep* and *cap* ORFs), with subsequent infection with wildtype Adenovirus (Samulski *et al.*, 1987). This method proved undesirable as eliminating wildtype Adenovirus post-production was problematic and laborious. The process also raised caution with respects to the safety profile of rAAV vectors produced in this manner. Subsequent development to this process involved the use of a two-plasmid system (as above); however, the helper plasmids included AAV *rep/cap* and Adenovirus supporting functions. This

ensured a helper virus-free system that resulted in an 80-fold increase in vector production compared to methods using wildtype Adenovirus. A two-plasmid system also ensured higher transfection efficiencies (Collaco *et al.*, 1999). Similarly, the design of a recombinant plasmid that separately provided helper functions (in that in the absence of Adenovirus co-infection) demonstrated higher levels of vector output than compared to production of AAV in the presence of Adenovirus co-infection (Matsushita *et al.*, 1998; Xiao *et al.*, 1998). Despite these milestones, investigations to scale-up the amount of rAAV vector from packaging and producing systems is on-going.

Developing production systems have now involved looking towards modifying cell lines in effort to improve upon the limited scalability of rAAV vectors. A number of modifications have been described in the literature, of which fall under one of four themes: i) generation of stable producer HeLa or A459 cell lines (Gao *et al.*, 2002b); ii) use of insect cell line Sf9 that is derived from ovarian tissue of *Spodoptera frugiperda* (Vaughn *et al.*, 1977); including an Sf9-based partial packaging cell line with stable expression of AAV *rep* and *cap* (Smith *et al.*, 2009; Mietzsch *et al.*, 2014); iii) use of HSV-1 systems (Kang *et al.*, 2009); and lastly, the most recent development, iv) manipulation of human cell lines and endogenous genes to improve vector production (Satkunanathan *et al.*, 2014; Satkunanathan *et al.*, 2017).

The relationship between cellular factors and rAAV vector production remains, to date, understudied. Whereas, in contrast there is substantial understanding of helper virus function on AAV and host cell biology. A number of reports do support a role of host cellular factors in AAV biology and production. For instance, it was shown that AAV vectors associated with a number of cellular proteins, including nucleophosmin (Dong *et al.*, 2014). Nucleophosmin functions mainly in the nucleolus, but studies between AAV2 and nucleophosmin indicated that it interacts with AAV Rep proteins (Bevington *et al.*, 2007), and functions to modulate particle mobilisation and genome uncoating (Johnson and Samulski, 2009). Therefore, Satkunanathan *et al.* (2014) aimed to decipher the cellular protein makeup that is either co-produced or co-purified with rAAV vectors *in vitro*, in further detail.

Using mass spectrometry Satkunanathan *et al.* (2014) identified several cellular proteins, including nucleophosmin, that were associated with rAAV vectors. Mass spectrometry identified the Y-Box binding protein (YB)1, which is a highly complex and multifunctional protein. It was suggested that YB1 may have an important function in rAAV vector production,

seeing as short-haripin (sh)RNA-mediated knockdown of YB1 in 293T cells exhibited an enhanced packaging cell line quality. A number of YB1 knockdown effects were documented: YB1 had a drastic effect on AAV biology with significant impact on both Rep and AAV2 genome turnover (increased vector production for AAV2 up to 45-fold increase). The effect of YB1 knockdown was profoundly most evident on rAAV2 serotype than compared to rAAV5 or rAAV8. This indicated that YB1 functioned in a serotype-specific manner. Taken together it seems that the molecular mechanisms of YB1 directly or indirectly impacted on rAAV vector production in a negative manner.

1.3.5 Recombinant baculovirus technology

Recombinant AAV vector production in Sf9 cell lines was adapted because of the ability for recombinant baculoviruses to infect insect cell lines and produce large quantities of heterologous proteins *in vitro*. This was due to the strong insect-specific promoters, such as p10 and polyhedrin (polh), used in baculovirus expression vector systems, which drive late acting, high level expression of protein (Graber *et al.*, 1992; Miranda *et al.*, 1997; Usami *et al.*, 2011; Wilde *et al.*, 2014). Additionally, insect cells such as Sf9 are easily grown in suspension culture and in serum-free conditions. Post-translational modifications of proteins synthesised in insect cells is an additional feature and benefit, as these processes occur (to a limited extent) similar as is seen in mammalian cells (Vrljic *et al.*, 2011). Given these main features, the production of heterologous protein or rAAV vectors could be produced in a scalable manner, especially when mediated by the use of recombinant baculoviruses.

Recombinant baculovirus technology is mainly derived from *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV), an insect-specific virus with a predominant infectivity for the larvae of the order Lepidoptera, which includes *S. frugiperda* and *Trichoplusia ni*. Towards the latter stages of AcMNPV infection, large amounts of polyhedrin gene expression was identified due to its promoter (polh) activity (Carstens *et al.*, 1979). However, the non-essential property of polyhedrin protein to virus replication in culture (Smith *et al.*, 1983a), drove the exploitation and utilisation of AcMNPV for baculovirus expression vector systems. Although, AcMNPV contains a very large circular genome (>130kb) (Ayres *et al.*, 1994), which proved difficult to manipulate using classical cloning strategies. Instead recombinant baculoviruses were generated by homologous recombination (HR) strategies upon co-transfection of insect cell lines between a transfer vector (containing gene of interest) and the non-essential *polyhedrin* gene region of AcMNPV circular genomic DNA. The *polyhedrin* gene acted as target for

recombination; however, considerable development to the baculovirus expression vector system for generation of recombinant baculovirus was required. The generation of recombinant baculovirus by the aforementioned method was a highly inefficient process, as only an average of 0.5% of progeny virus were evaluated as recombinant (Smith *et al.*, 1983b). This meant that highly time-consuming purification protocols were necessary to isolate recombinant baculovirus from wildtype.

The baculovirus expression system was further improved to diminish the inefficiency in recombinant baculovirus generation. This improvement mainly originated from attempts to make deficient the wildtype progeny. Strategies included linearization of the AcMNPV genomic DNA at either the *polyhedrin* or *p10* loci, through careful and targeted insertion of restriction enzyme sites (Kitts *et al.*, 1990). In theory, using linearised baculovirus genomic DNA for recombination with an appropriate transfer vector would permit only transfection of re-circularised genomic DNA, and therefore generation of recombinant baculovirus populations only. This approach increased the fraction of recombinant baculovirus from wildtype to approximately 30% (Kitts *et al.*, 1990), which in itself was a substantial improvement from the near impractical 0.5% efficiency.

Soon after, this system was further improved to permit highly efficient recombinant baculovirus production with up to 98% efficiency (Kitts and Possee, 1993). Essentially, the adapted production process resulted in further hampering the ability for wildtype progeny from being produced. This adapted system took advantage of the essential gene, *orf1629*, found just downstream of the *polyhedrin* locus. The *orf1629* was truncated in the engineered baculovirus genomic DNA, coupled with the linearisation at the *polyhedrin* locus. Instead the remainder of the *orf1629* was found in the transfer vector used to produce recombinant baculovirus upon co-transfection. This meant that upon co-transfection, successful recombination between linearised baculovirus genomic DNA and transfer vector would restore essential *orf1629* gene function, and permitted only recombinant baculovirus progeny to be produced.

Further innovative strategies to produce clonal recombinant baculoviruses in a direct, and almost one-step manner were developed soon after. This allowed for quick production of recombinant baculovirus, which does not necessarily require purification by sequential plaque assays. The technology and production method was developed so that all baculovirus genomic

DNA product contained the heterologous gene of interest. Such methods involved production of recombinant baculovirus by site-specific transposition events between the genomic DNA of baculovirus (engineered into a bacterial artificial chromosome [BAC]), and shuttle vector using bacteria systems. Instead, in bacterial cells the process was simplified and proved more efficient as the entire baculovirus genomic DNA was recombined into a BAC at the *polh* locus. This allowed the engineered BAC DNA to propagate in *E. coli* (Luckow *et al.*, 1993; Zhu and Qi, 1999). Shuttle vectors in this context were referred to as bacmids. An additional advantage of this methodology meant that transformation of *E. coli* harbouring the engineered BAC with desired bacmid would allow for site-specific transposition between the Tn7 transposon elements present in the baculovirus portion of the BAC and bacmid. Ultimately, the developed system allowed for propagation of a single genotype of recombinant baculovirus genomic DNA post site-specific transposition. The resultant genome can subsequently be used to transfect insect cells for recombinant baculovirus production, in a more than less clonal manner than previously described. Resultant recombinant baculoviruses were referred to as (r)baculovirus expression vectors (BEVs).

1.3.6 Recombinant baculovirus technology for rAAV vector production

Virus-like particles (VLPs) have successfully been generated using the baculovirus expression system, and includes VLPs derived from parvovirus B19, human parvovirus, and even AAV (Ruffing *et al.*, 1992; Tsao *et al.*, 1996; Aucoin *et al.*, 2008; Abdoli *et al.*, 2013). In the case of AAV VLPs, the AAV2 Vp1, Vp2, and Vp3 sequences were individually expressed or co-expressed from single recombinant baculoviruses. Upon infection of Sf9 cells, expressed Vp1-3 showed competency for VLP formation, in that empty AAV vectors were synthesised. However, the formation of AAV2 VLPs was limited to situations where Vp2 expression was present and Vp2 seemingly promoted capsid assembly in its entirety. Single or combination infection with recombinant baculoviruses expressing Vp1 or Vp3 did not yield detectable formation of VLPs (Ruffing *et al.*, 1992). Nonetheless, initial proof of principle was obtained that AAV particles could be produced in Sf9 cells.

The use of recombinant baculovirus and Sf9 cells for rAAV vector production was further exploited by Urabe *et al.* (2002), and involved the expression of AAV2 *rep* and *cap* ORFs and AAV2 *ITR*-containing vector (GFP expressing vector) from three separate recombinant baculoviruses. AAV2 Rep78 and Rep52 encoding sequences were split and cloned into a single bacmid, and controlled under the truncated *immediate-early 1* gene promoter, Δ IE1 (derived

from *Orgyia pseudotsugata* [Oc]MNPV), and p10 respectively. This differential use of promoters to drive Rep78 and Rep52 protein expression separately allowed optimal protein expression dynamics for rAAV vector production, as it was more previously noted that unregulated overexpression of Rep78 and Rep68 inhibited rAAV titres (Li *et al.*, 1997). Although, these observations were concluded from studies involving Adenovirus-infected or Adenovirus-helper transfected mammalian cells as opposed to in an insect cell line. Nonetheless, AAV2 *cap* ORF was cloned into a separate bacmid under the polh promoter control. Additional adjustments were also made to ensure the near expected 1:1:10 stoichiometry of Vp1-3 expression by mutating the *cap* ORF start codon (for Vp1) from ATG to ACG, and a 9nt sequence element found upstream to Vp2 ACG start codon was added in a similar position relative to Vp1 start codon (Urabe *et al.*, 2002). The *ITR*-containing bacmid, harbouring the GFP transgene, was also modified such that the GFP transgene was cloned under the control of both CMV and p10 promoters. This permitted the resultant baculovirus (BEV-GFP or BEV-vector) to permit expression in both insect cell lines and mammalian cell lines. Ultimately, the *ITR*-containing bacmid contained the necessary *cis*-acting elements required for AAV genome replication and packaging. The resultant system meant that Sf9 producer cells were co-infected with three separate baculoviruses in order to instigate rAAV2 vector production (Fig. 1.9).

Though the system of using three recombinant baculoviruses to produce rAAV vectors from Sf9 cells by Urabe *et al.* (2002) showed promising, some drawbacks of the adapted technology proved it limiting. For instance, having to use three different recombinant baculovirus to infect Sf9 cells for rAAV vector production relied on each baculovirus to infect a single Sf9 cell for vector production. This was in part offset by the productive infection by recombinant baculovirus – in that baculovirus infection of insect cell lines results in the generation of progeny recombinant baculovirus harbouring the heterologous gene. These are able to bud out of the initial infected cell and superinfect the same cell or neighbouring cells (Xu *et al.*, 2013). Although, it is noted that high MOI's are required to infect insect cells for AAV vector production, meaning high titre stocks of rBEVs are usually required. Another drawback was found to be genetic instability of the rBEVs. Scale-up of rBEVs is an advantageous feature of using the technology for heterologous protein expression in insect cell lines. However, it has been repeatedly observed that genetic instability of recombinant baculovirus was a feature from vectors derived from the bacmid system. In fact, serial passage of the recombinant baculovirus in insect cell culture resulted in eventual loss of the transgene or even substantial

losses in the baculovirus genomic DNA (Pijlman *et al.*, 2003; Kohlbrenner *et al.*, 2005; Negrete *et al.*, 2007). Concerning the recombinant baculovirus encoding the *rep78* and *rep52*, the sequence that encodes Rep52 was also present in the sequence that encodes Rep78. A high degree of homology is shared between the two transgenes. Furthermore, given the orientation in which these transgenes were cloned a perfect palindrome of the homologous sequences within the same baculovirus vector DNA was present. The sequences within this particular vector were inclined to show genetic instability and HR, which led to the loss of the transgenes upon serial passage.

To circumvent the loss of the *rep* transgenes, the AAV2 *rep* ORF was modified to permit the expression of Rep78 and Rep52 from a single mRNA transcript (Smith *et al.*, 2009). This was achieved by exploiting the insect cell line's ribosome scanning mechanism. The *rep* ORF (under the promoter control of *polh*) was modified, predominantly at the *rep78* initiation codon, from AUG to CTG, to allow suboptimal initiation of translation. On the other hand, the ribosome scanning property allowed optimal expression of Rep52 due the presence of a suitable AUG initiation codon by leaky scanning mechanisms. Additionally, in order to simplify the rAAV vector production system, the *cap* ORF (driven by p10 promoter) was cloned into the same bacmid vector in the opposite transcriptional orientation. This allowed production of only two different recombinant baculoviruses for rAAV vector production in Sf9: i) the *ITR*-containing transfer vector (BEV-vector), and ii) the dual *rep* and *cap* encoding vector (BEV-RepCap), which determined the serotype. The end result was a simplified system to produce rAAV vector from Sf9 cells using two rBEVs (Fig. 1.9), as well as, the genetic stability of the rBEV-RepCap showed significant improvement (Smith *et al.*, 2009).

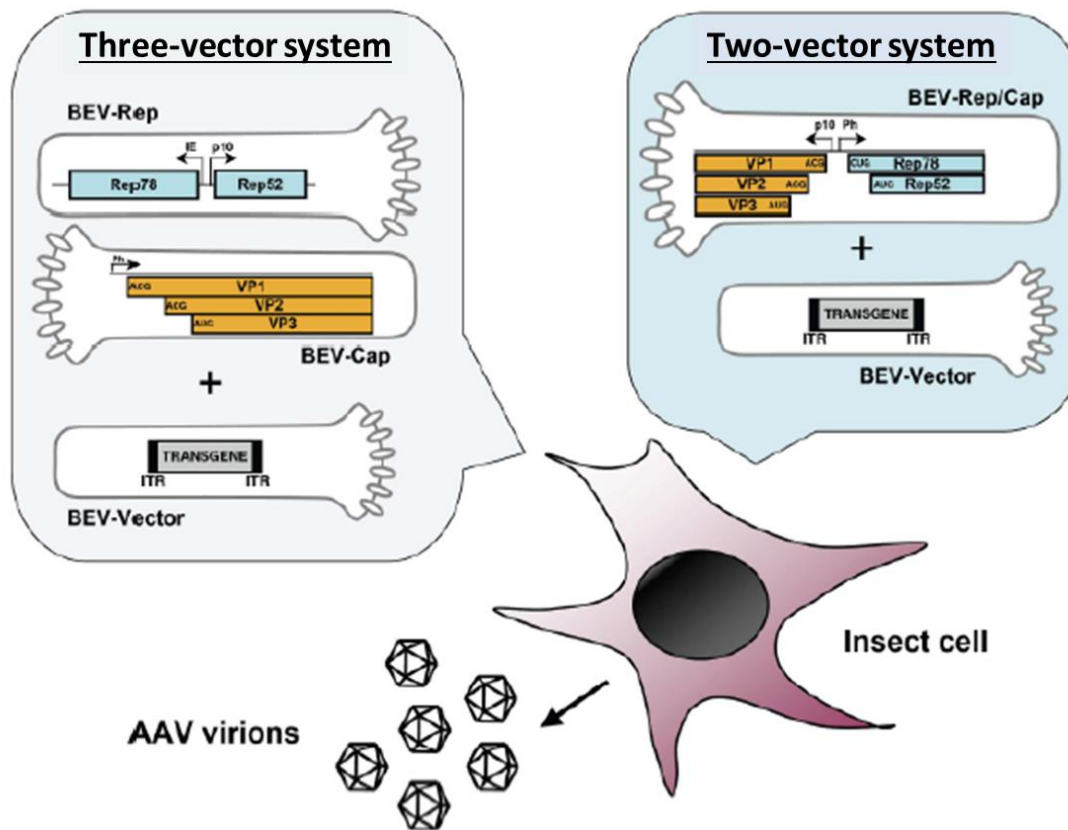


Figure 1.9 Recombinant AAV vector production using recombinant baculovirus technology. Insect cells, typically Sf9 cells, are used to generate rAAV vectors. In particular, production in insect cells does not require adenoviral-derived helper functions, as these are provided by the baculovirus infection and use of insect-specific promoters. Ultimately, two main production systems for rAAV vector in Sf9 cells are described – the three-vector and two-vector systems. The former three-vector system involves three separate BEVs (BEV-Rep, BEV-Cap, and BEV-vector) to infect Sf9 cells and produce rAAV vector of a desired serotype. The two-vector system includes two separate BEVs - BEV-Rep/Cap, which includes both *cap* and *rep* ORFs, and BEV-vector, which supplies the AAV genome and transgene. Figure adapted from Ayuso *et al.* (2010).

It is important to note that unlike production systems using mammalian cell lines, Adenovirus-derived helper functions supplied in *trans* are entirely unnecessary for rAAV vector production using insect cell lines and baculovirus technology. This is because the natural promoters used to drive expression of *rep* ORF (p5) and *cap* ORF (p40) were replaced with the aforementioned insect-specific promoters, which would be activated and upregulated upon infection of Sf9 cell line with the relevant baculoviruses (Smith *et al.*, 2009). In essence, the infection of Sf9 cells by the recombinant baculovirus provides the sole helper function required to drive p10, polh, or Δ IE1 promoter activities, and in turn expression of AAV Rep and Cap proteins for rAAV vector assembly and vector genome amplification.

1.3.7 Identification of human YB1 as a putative AAV restriction factor

Little direct association between human YB1 and AAV biology has been identified, except by

Nash *et al.* (2009) who found that human YB1 co-immunoprecipitated with AAV2's non-structural protein Rep78, and by Satkunanathan *et al.* (2014) using shRNA-mediated knockdown of YB1 in 293T cells. YB1 co-immunoprecipitation of Rep78, including additional others explored by the study. It was implied, given YB1's characterised functions, that the expression of AAV proteins under Rep78 and p5, p19, p40 promoter may be affected by YB1 binding (Nash *et al.*, 2009). The effect of this binding on AAV biology, however, was not further explored. However, LC-MS/MS identified YB1 associated with wildtype AAV (Nash *et al.*, 2009) and rAAV vectors (Satkunanathan *et al.*, 2014), which prompted further characterisation of YB1's impact on rAAV vector biology and output. In fact, performing YB1 knockdown studies correlated with enhanced rAAV2 and rAAV8 vector titres *in vitro*. Therefore, YB1 potentially functions negatively with respect to AAV in a serotype-specific manner. Importantly, YB1 knockdown was associated with enhanced AAV Rep protein expression, vector genome copies, and improved full:empty particle ratios after triple transfection for rAAV vector production (Satkunanathan *et al.*, 2014). Therefore, an enhanced AAV vector producer cell line was established, which showed an impressive up to 1.5 log increase in rAAV2 or 8 vector titres, and it was suggested that YB1 seemed to possess antiviral properties that limit AAV virus production and replication (Satkunanathan *et al.*, 2014). We next explore the functions and properties of human YB1 in section 1.4.0, below.

1.4.0 Cellular protein YB1

1.4.1 Introduction to YB1 and its structure

Bacteria have evolved the ability to respond to a number of environmental stresses, including a rapid decrease in temperature – 'cold shock'. Bacteria respond to a 'cold shock' by shutting down molecular and cellular machineries, including reducing the efficiencies in transcription and translation (Nakaminami *et al.*, 2006). A means to overcome this environmental stress and prevent stress-induced death is mediated by cold shock proteins, which function to permit bacteria to adapt and survive these extreme environmental changes.

Y-Box proteins encompass a highly evolutionary conserved family of nucleic-acid binding proteins between prokaryotes and eukaryotes (Karlson and Imai, 2003; Nakaminami *et al.*, 2006). Y-Box proteins harbour the cold shock domain (CSD), which is homologous to cold shock proteins (Mani *et al.*, 2012; Kljashtorny *et al.*, 2015), and represents the most conserved, defining feature of Y-Box proteins. Human YB1 protein is better described as a highly complex protein with great number of functions within the cell. Not only does YB1

possess DNA/RNA-binding functions, YB1 is also capable of binding to proteins, making the protein highly unconstrained in function. YB1 was first recognised as a component of cytoplasmic messenger ribonucleoprotein (mRNP) in the 1970's (Kumar and Pederson, 1975). However, it was so aptly named for its then preferential binding to a DNA nucleotide sequence called the Y/CCAAT box (Y-box) – 5'-CTGGATTGG C/T C/T AA-3' (Didier *et al.*, 1988).

YB1's functions are dictated by its ability to form homo-complexes, complexes with other unrelated proteins, and its interactions with nucleic acids. YB1 is also widely expressed throughout development, and that YB1 deficiency has been predicted to be incompatible for cell survival based on transgenic mice models (Lu *et al.*, 2005; Uchiumi *et al.*, 2006). Transgenic mice models were developed for YB1 knockout, but YB1^{-/-} showed embryonic lethality. Some embryos did manage to develop relatively normally at early stage development, but during late development YB1 was deemed more essential and non-redundant in function (Lu *et al.*, 2005).

Human YB1 is encoded by the *YBX1* gene, found on chromosome 1 (Ch1p34.2), and the gene itself is composed of 8 exons that span across approximately 19kbp. YB1 protein is a 324aa long, 36kDa protein and is composed of three structural domains. The N-terminal domain is the alanine/proline-rich (A/P) domain (1-50aa). Next is the highly evolutionary conserved CSD which covers 51-129aa. Lastly, the C-terminal domain (CTD) encompasses the rest of YB1, and is formed of alternating clusters of positively and negatively charged residues. A diagrammatic representation of YB1 and its domains is depicted in Fig. 1.10.

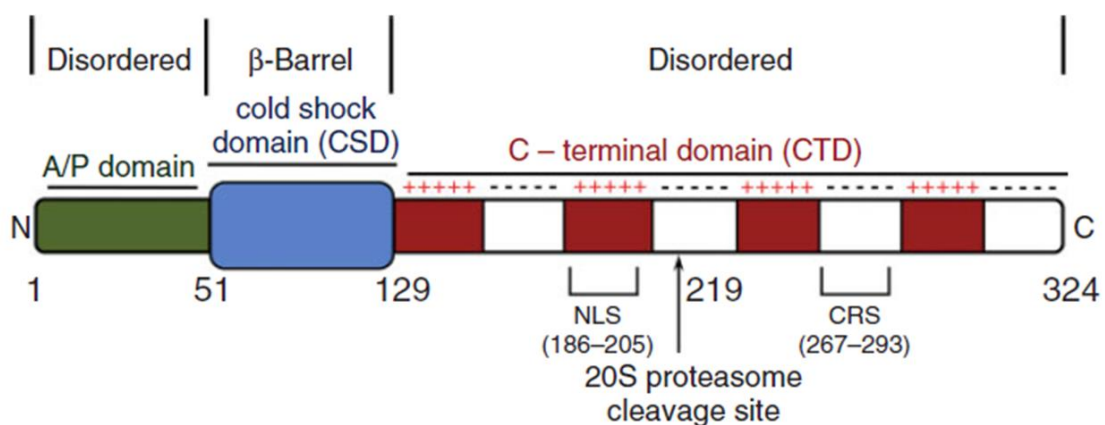


Figure 1.10 Organisation of YB1 protein. YB1 protein is composed of three structural domains: A/P domain (1-50), CSD (51-129), and CTD (130-324). The A/P domain is so called because of its richness in alanine and proline residues. The CTD domain contains alternating clusters and positively (+; regions coloured red) and negatively (-; white coloured regions) charged residues. Both A/P and CTD domains are considered highly disordered. A NLS is found at position 186-205, and a CRS is found at position 267-293. A 20S proteasome cleavage site at position 219 is also annotated. Figure from Lyabin *et al.* (2014).

The A/P and the CTD domains do not appear to form definitive or regular secondary structures. Instead they are predicted as intrinsically disordered because these domains have yet to be solved for their structure in solution (Guryanov *et al.*, 2012). The presence of type II polyproline helical structures dictates intermolecular interactions between YB1 and other proteins (Guryanov *et al.*, 2013). The CTD is responsible for most protein-protein interactions, of which alternative basic and acidic aa residues can potentially bind to RNP complexes, and therefore shuttles them between the nucleus and cytoplasm (Raffetseder *et al.*, 2003). This shuttling property is facilitated by the presence of a NLS and cytoplasmic retention signals (CRS), which are mainly found in the CTD (Fig. 1.10).

On the other hand, the solution structure of YB1's CSD has been determined and modelled (Kloks *et al.*, 2002), and its tertiary structure resembles that of bacterial cold shock proteins (Wang *et al.*, 2000b; Kljashtorny *et al.*, 2015). The CSD is a highly conserved domain, which adopts a five-stranded β -barrel tertiary structure. The β -strands are arranged in an anti-parallel manner and connected by loops. The second and third strand of the β -barrel follows a consensus sequence of K/N-G-F/Y-G-F-I/V and V-F-V-H-F (Landsman, 1992), otherwise known as the RNP-I and II motifs, respectively (Kloks *et al.*, 2002). YB1's ability to bind to DNA and RNA is mediated by these consensus motifs, but binding to dsDNA, ssDNA and RNA species (Izumi *et al.*, 2001; Zasedateleva *et al.*, 2002; Kljashtorny *et al.*, 2015) is relatively non-specific; or rather, not confined to the Y-box. In fact, YB1 showed preferential binding to ss GGGG motifs and G-rich RNA, and by comparison binds with less affinity to Y-Box motifs as it was originally named for (Zasedateleva *et al.*, 2002; Kljashtorny *et al.*, 2015).

1.4.2 YB1's role in transcriptional and translational regulation

YB1's function in transcriptional regulation was initially proposed due to YB1's ability to bind to DNA Y-Box sequence in promoter regions of the major histocompatibility complex II gene, *HLA-DR α* (Didier *et al.*, 1988). Its role as a potential transcription factor was supported further by studies which showed its preferential binding to ssDNA sequences compared to dsDNA (Izumi *et al.*, 2001), and that it possesses the abilities to melt or anneal nucleic acids (Skabkin *et al.*, 2001). It is now better established that YB1 is not limited to binding to just the Y-Box motif. This flexibility allows YB1 to upregulate or downregulate a number of gene targets, including those involved in cell proliferation and division, cell cycle arrest, apoptosis, and multiple drug resistance, and its role in these cellular processes will be outlined in brief in following sections. For example, in response to IFN γ , YB1 has been found to associate with an

IFN γ -response element (that is present in *COL1A2* gene promoter sequence). In response to IFN γ treatment YB1 repressed the transcription of the *COL1A2* gene. Interestingly, a Y-Box sequence is also present in *COL1A2*'s gene promoter, but transcriptional regulation via YB1 was predominantly mediated via the IFN γ -response element (Higashi *et al.*, 2003). The IFN γ -response element coincides with binding sites for transactivators Sp1 and Sp3, and therefore, it is possible that the sequence region is under competitive control between YB1 and relevant transactivators to regulate *COLA2* gene transcription.

Transcriptional regulation by YB1 is not limited to transactivation or repression of genes. It has been shown to be involved in regulating pre-mRNA splicing and alternative splicing events, as well. It is a component of the spliceosome (Hartmuth *et al.*, 2002; Deckert *et al.*, 2006), which functions to mediate splicing events to generate mRNA fit for translation into protein. Although the exact function of YB1 in pre-mRNA splicing has been challenging to elucidate, few studies have shown its direct involvement as a splicing activator when concerning exons/introns that harbour A/C-rich exon enhancer (ACE) elements (Stickeler *et al.*, 2001). YB1 was associated with increased activation of exon inclusion of the *CD44* exon v4, and this was dependent on the presence of ACEs (Stickeler *et al.*, 2001); the binding of which promoted inclusion of exon v4 into the *CD44* mRNA product. The alternative splicing of *CD44* exon v4 is further regulated by YB1's association with the PP2C γ protein (a Ser/Thr phosphatase); of which involvement in splicing is relatively unknown but at the very least promotes the assembly of the spliceosome. However, its phosphorylation enables its binding with YB1, and knockdown of PP2C γ compromises splicing of *CD44* pre-mRNA in context with exon v4 (Allemand *et al.*, 2007). Direct interaction of YB1 with the repressor of 3' splice site utilisation, SRp30c, was also demonstrated. This interaction skewed the preferential splicing pattern of the Adenovirus *E1A* mini-gene (Chansky *et al.*, 2001; Raffetseder *et al.*, 2003). More recently, YB1 facilitated the activation of alternative splicing and exon inclusion by binding the CAUC and CACC motifs in target exon and upstream intron, to recruit U2AF65 splicing factor and in turn mediate the exon inclusion of *CD44* exon v5 (Wei *et al.*, 2012).

YB1 was identified in eukaryotic cells as a major protein factor that helps to make up polyribosomal mRNPs, and predominantly helps make up inactive, non-polyribosome-bound mRNPs (Kumar and Pederson, 1975; Minich *et al.*, 1993). The exact mechanism in which YB1 functions to regulate translation seems complex, but the protein generally exhibits properties that inhibit translation or help stabilise mRNA. YB1 was shown to negatively affect the

assembly of 48S pre-initiation complex to β -globin mRNA using an *in vitro*, cell-free translation system (Pisarev *et al.*, 2002). Low levels of cytoplasmic YB1 relative to mRNA correlates with active translation and protein synthesis. However, with high levels of YB1 relative to mRNA, translation initiation is repressed (Davydova *et al.*, 1997; Pisarev *et al.*, 2002). With a low YB1/mRNA ratio, YB1 binds to mRNA as monomer units along the length of mRNA using its CSD and CTD, but homo-oligomeric YB1 complexes associate with mRNA instead when YB1/mRNA ratio is high. Multimerisation occurs through YB1's CTD, and the CSD facilitates binding to mRNA. The transition may promote free mRNP formation, and prevent translation factors associating with the mRNP because excess YB1 packages mRNA so that its termini are thought to be buried inside its complex, preventing the key components of translation from assembling (Skabkin *et al.*, 2004).

YB1's CSD was found to be the main functional domain that facilitated mRNA stability in a 5'-GTP or m⁷GTP cap-dependent manner. In this way it protects mRNA from degradation (Evdokimova *et al.*, 2001). The 5'-cap structure is well known in permitting mRNA stability and allowing mature mRNA to undergo translation efficiently (Gillian-Daniel *et al.*, 1998). Therefore, uncapped mRNAs would be susceptible to degradation. Evdokimova *et al.* (2001) found that YB1 sequesters to 5'cap to improve mRNA stability and regulate translation, predominantly by YB1's CSD. Its CTD was reportedly responsible for the inhibition of translation initiation by interfering with eIF4G binding to mRNA, and preventing the initiation of translation (Nekrasov *et al.*, 2003). This could be a complementary effect to promote mRNA protection, especially in context with mRNAs and mRNPs that are not bound to ribosomes and are translationally inactive. Therefore, during active translation, YB1 levels are 'optimal' to permit accessibility of translation initiation factors to bind to mRNA. Displacement of initiation factors by YB1 from the length of the mRNA promotes the accumulation of relevant initiation factors at the 5'-cap end, which in turn allows the assembly of functional translational machineries.

1.4.3 YB1's function in cell proliferation and the cell cycle

YB1 has been frequently linked to tumour malignancy, and its upregulation is commonly seen in a number of solid tumours, including: breast, lung, prostate cancers (Gimenez-Bonafe *et al.*, 2004; Fujita *et al.*, 2005; Kashihara *et al.*, 2009) and melanoma (Sinnberg *et al.*, 2012). YB1 has been argued to have significant prognostic value. As a prognostic marker, aberrant YB1 expression in certain tumours has been associated with worse prognosis (Dahl *et al.*, 2009;

Fujiwara-Okada *et al.*, 2013). YB1's role in cell proliferation was initially postulated due to phenotypic changes that resulted from targeted disruption of the Chicken-YB1 gene in DT-40 cells (Swamynathan *et al.*, 2002). Deletion of the 5' end of one of the Chicken-YB1 alleles led to cell cycle defects, with slowed cell proliferation, and increased apoptosis rates. Additionally, silencing YB1 in SH-SY5Y neuroblastoma cell line using shRNA resulted in G₀/G₁ arrest with suppressed cell growth rates and an increased susceptibility to apoptosis (Wang *et al.*, 2015b). A benefit of the study by Wang *et al.* (2015b) was the *in vivo* examination of YB1's role in cell proliferation using xenograft tumour models. Silencing YB1 expression in SH-SY5Y cells and inoculating these subcutaneously in nude mice resulted in slower growing tumours, while intra-tumour injection of SH-SY5Y-derived tumours with shRNA plasmid targeting YB1 was associated with reduced tumour sizes.

YB1, therefore, exerts a positive effect on cell proliferation and the cell cycle, and currently, this seems to be partly due to a relationship between YB1 and cyclins (a family of protein involved in the regulation of cell cycle progression). YB1 was shown to stimulate *CCND1* gene expression by binding to its promoter and upregulating its expression, whereas when YB1 was silenced, *CCND1* gene expression was repressed, resulting in a stall in cell cycle progression and proliferation (Fujiwara-Okada *et al.*, 2013; Wang *et al.*, 2015b). However, it was appreciated that most studies that explore the role of YB1 in relation to cell proliferation and cell cycle progression, chiefly make use of cancer or immortalised cell lines and not primary cell lines to postulate the role of YB1. Work on human and mouse primary cell lines provided evidence that YB1 directly regulates the expression of the tumour suppressor gene, *p16*, at the *INK4* locus (Kotake *et al.*, 2013). Protein p16 is a cyclin-dependent kinase inhibitor and negatively regulates cell cycle progression by inhibiting progression into S-phase. Upregulated expression of YB1 correlated with reduced *p16* gene expression, and as a result primary cells were not as easily succumbing to cellular senescence.

1.4.4 YB1's function in DNA repair

YB1's suggested role in DNA repair was postulated based on YB1's increased binding affinity for apurinic DNA (Hasegawa *et al.*, 1991). This was supplemented by the fact that YB1 harbours 3'-5' exonuclease and endonuclease activities (Izumi *et al.*, 2001; Gaudreault *et al.*, 2004). Additionally, use of heterozygous mouse embryonic stem cells (YB1^{+/-}) implied YB1's involvement in repair of DNA damage or adducts induced by DNA crosslinking agents, including: mitomycin C and cisplatin. Treatment of YB1^{+/-} mouse embryonic stem cells with

these agents showed increased sensitivity and cytotoxicity, especially when compared to the wildtype YB1^{+/+} control (Shibahara *et al.*, 2004). Furthermore, YB1 was successfully co-immunoprecipitated with Mre11 and Rad50 proteins, which comprise the DNA repair complex, MRN (Mre11-Rad50-Nbs) complex. This co-immunoprecipitation was only evident from NIH3T3 cell lysates that had been treated with etoposide or doxorubicin, and absent from lysates derived from non-treated NIH3T3 cells, both transiently expressing a truncated YB1 (Kim *et al.*, 2013). There was no indication that exogenously expressed YB1 (full length) could co-immunoprecipitate the same proteins. Association with additional protein factors of DNA repair pathways had been evidenced more previously by Gaudreault *et al.* (2004). Here, instead, GST-YB1 fusion protein was immobilised and 293 cell extract was loaded to demonstrate that full length YB1, albeit without DNA damaging stimuli, could bind additional proteins involved in DNA repair; namely, MSH2, DNA polymerase δ , Ku80 and WRN proteins. Collectively, these studies imply that YB1 may have functions in base-excision and mismatch repair pathways *in vitro*.

The above studies do not in its entirety directly show YB1 has a direct function in DNA repair pathways. The evidences simply imply a role for YB1 based on correlative association and/or binding with DNA repair proteins *in vitro*, and does not illustrate the exact significance of the associations. However, Das *et al.* (2007) firstly identified YB1's physical association with Neil-like-2 (NEIL2), and its functional complex with DNA polymerase β and ligase III α ; but additionally provided evidence that the YB1 interaction with NEIL2 resulted in enhanced enzymatic activity after oxidative stress *in vitro* (Das *et al.*, 2007). NEIL2 is a key player in base-excision repair of DNA by functioning as a DNA glycosylase and initiating the DNA repair pathway. Therefore, YB1's involvement in base-excision repair was partly elucidated *in vitro*, such that the presence of YB1 enhances the enzymatic activity of NEIL2. Interestingly, YB1's role in DNA repair is not limited to reparation of chromosomal DNA, but also seems to be involved in maintaining the integrity of mitochondrial DNA. YB1 was identified as capable of binding to mismatched DNA, and to DNA mismatch repair proteins present in the mitochondria (de Souza-Pinto *et al.*, 2009). Using MMR assays (M13 assays involving circular substrate with a mismatch in the β -galactosidase encoding complementary [c]DNA) with mitochondrial extracts from control and YB1 depleted cell lines, a significantly reduced degree of repair activity was observed when YB1 was deficient (de Souza-Pinto *et al.*, 2009). Therefore, YB1 has been consistently observed to be associated with nuclear and now mitochondrial DNA repair pathways, and the deficit leads to inefficient DNA repair *in vitro*.

1.4.5 YB1's function in cellular stress

YB1 predominantly exists in the cytoplasm, and in response to cell stress cytoplasmic YB1 functions to repress translation initiation to promote cell survival. Translation initiation is put on pause by phosphorylation of translation initiation factor eIF2 α upon the detection of certain cell stresses including hypoxia, oxidative stress, heat shock, to name a few (Pavitt *et al.*, 1998). As a result, stalled translation preinitiation complexes organise themselves in the cytoplasm and form stress granules. These are composed of non-ribosome-bound mRNAs, translation initiation factors, and YB1, and represent a form of mRNA protection and storage that are translationally inhibited during the duration of adverse cell conditions (Yang and Bloch, 2007).

Despite the close association of YB1 with stress granules, the exact function of YB1 in stress granule formation or in stalling mRNA translation remains to be fully characterised. Studies exploring YB1's role demonstrate that YB1 regulates the post-transcriptional expression of *G3BP1* mRNA; the product of which is a nucleating factor required for stress granule formation (Somasekharan *et al.*, 2015). It was identified in U2OS cells that YB1 binds directly to *G3BP1* mRNA, presumably within its 5'-untranslated region (UTR) sequence to instigate the translation of the *G3BP1* mRNA, and allow the formation of stress granules under the stress conditions tested. In contrast to this, the downregulation of YB1 was associated with reduced *G3BP1* mRNA translation, and compromised stress granule formation. However, YB1 knockout cell lines generated from U2OS cells by Lyons *et al.* (2016) were also defunct in stress granule formation. But this was unrelated to *G3BP1* expression. Despite the knockout of YB1, *G3BP1* expression remained unchanged, thus implying that stress granule formation is regulated by YB1 in a more stress-specific manner.

Alternatively, tRNA cleavage products, 5'-tiRNA^{Ala} and 5'-tiRNA^{Cys}, occur as a result of angiogenin RNase activation in response to cell stress (Li and Hu, 2012). These tiRNAs are able to induce the formation of stress granules and inhibit translation initiation in a phospho-eIF2 α -independent manner. Additionally, tiRNAs have been shown to bind to YB1 via its CSD (Ivanov *et al.*, 2011; Ivanov *et al.*, 2014). These tiRNAs function by displacing the eIF4F from m⁷GTP-capped mRNA and prevents translation initiation. This process seemingly requires YB1 to efficiently mediate this process as determined by YB1 knockdown studies (Ivanov *et al.*, 2011). However, cell free translation systems using YB1 knockout U2OS cell lysates showed that tiRNAs were able to maintain their ability to repress translation initiation independent of YB1

(Lyons *et al.*, 2016).

Tumour progression can present with multidrug-resistance to chemotherapies. Typical chemotherapies are cytotoxic agents that aim to destroy cancerous cells; however, the eventual development of multidrug-resistance circumvents this. Multidrug-resistance is associated with the overexpression of P-glycoprotein, encoded by the multidrug-resistance 1 (*MDR1*) gene (Trock *et al.*, 1997). P-glycoprotein functions as an efflux pump that actively pumps out various substrates, including anticancer drugs, from cells (Trock *et al.*, 1997). This poses a significant issue when treating certain cancers. *MDR1* gene expression seems to be controlled by YB1, because YB1 was found to bind to a previously identified Y-Box sequence present on the *MDR1* promoter (Goldsmith *et al.*, 1993). Transcriptional activation of the *MDR1* gene correlated with YB1 levels and/or nuclear translocation only in cells treated by certain stresses, including UV irradiation or genotoxic agents (Ohga *et al.*, 1998; Shibahara *et al.*, 2004). Although, YB1 modulation of *MDR1* expression and chemoresistance is not a feature of all solid tumours; for example: PC-3, LNCaP and 22Rv1 cell lines derived from prostate cancer have shown that YB1 overexpression had no effect on *MDR1* expression when treated with chemotherapeutic agent docetaxel (Saupe *et al.*, 2015).

1.4.6 YB1's function in apoptosis

We have examined YB1's function in cell proliferation (Section 1.4.3) and in response to cell stress (Section 1.4.5). It comes to no real surprise that YB1 is also implicated in functioning in or regulating apoptosis. Transcriptional targets of YB1 include genes that are associated with cell death, including the Fas-receptor, *FAS* (Lasham *et al.*, 2000). This initiates pro-apoptotic signalling upon its binding to its cognate ligand. A number of studies also implicate YB1 and the tumour-suppressor, p53, which shows direct interactions (Okamoto *et al.*, 2000). YB1 was found to repress the *TP53* gene promoter, which was associated with the downregulation of *TP53* gene expression (Lasham *et al.*, 2003). Moreover, inducing *TP53* expression and activating p53 in C/ERp53/7 cells resulted in the translocation of YB1 into the nucleus and cell cycle arrest. YB1 is a dominant negative regulator of p53, and functions by binding to active p53 to inhibit its binding to promoter regions and transactivate genes *BAX*, *NOVA* (Homer *et al.*, 2005), and *APAF-1* (Zhang *et al.*, 2003). *BAX* and *NOVA*, which encode pro-apoptotic factors Bax and Nova, respectively, are therefore not upregulated to contribute to cellular apoptosis when YB1 is in functional interplay with activated p53.

1.4.7 YB1 and viral biology

YB1 has been reported important to select viral life-cycles. Human immunodeficiency virus (HIV)1 genomic and transcribed RNA species are prone to the host cells' RNA quality systems. RNA viruses have evolved measures to permit its longevity within infected cells. One way this is achieved is by protecting its RNA from both degradation and recognition by host RNA quality systems. HIV1 is capable of hijacking YB1 to stabilise viral RNA. As a result this led to enhanced HIV1 gene expression and virus production (Mu *et al.*, 2013). This was mainly identified by co-transfecting 293T cells with a plasmid expressing YB1 as well as HIV1 constructs, and measured luciferase activity was increased in YB1 overexpressing producer cells. Downregulation of endogenous YB1 by siRNA correlated with reduced luciferase activity in producer cell lines and reduced viral-vector titres (Mu *et al.*, 2013). RNA decay was significantly more stable in co-transfected 293T cells overexpressing YB1 (Mu *et al.*, 2013). This contradicts conclusions made by Ansari *et al.* (1999), who demonstrated that interaction of YB1 with HIV1 TAR RNA sequence and Tat protein correlated with increased Tat-induced and basal transcriptional activity as opposed to RNA stability. An enhanced transcriptional profile was not observed by Mu *et al.* (2013); however, increased RNA could be attributed by either more stable RNA forms or enhanced transcriptional activity. Nonetheless, YB1 is utilised by HIV1 in order to increase RNA levels to promote production. Similar was reported by Li *et al.* (2012) with respects to murine leukaemia virus (MLV) – MLV genomic RNA was stabilised by YB1 binding to the repeat region of MLV genomic RNA. This allowed for more available genomic RNA to be packaged for higher titres of recombinant MLV.

The mouse mammary tumour virus (MMTV) is a murine β -retrovirus, which instead seems to co-opt YB1 via Gag in an RNA-dependent manner. YB1 associates with genomic MMTV RNA during the formation of immature capsids in cytoplasmic granules. It was suggested that YB1 sequesters viral RNA away from translation machineries, and helps localise Gag to viral RNP (vRNP) cytoplasmic foci, which ultimately promotes virion assembly (Bann *et al.*, 2014). Influenza A virus also has shown to be regulated by YB1 (Kawaguchi *et al.*, 2012). Evidence suggests that YB1 is hijacked by Influenza A virus, relocates to the nucleus, and associates with promyelocytic leukaemia (PML) nuclear bodies in infected cells. Here YB1 complexes with Influenza A virus vRNPs in the PML nuclear bodies in order to facilitate their export from the nucleus, for further downstream trafficking via microtubules (Kawaguchi *et al.*, 2012; Kawaguchi *et al.*, 2015a). YB1-overexpressing cells correlated with higher virus titres, despite no significant change in the expression of viral proteins or viral RNA levels (Kawaguchi *et al.*,

2012). This increase in titres was only valid in context of Rab11a GTPase protein and its association with positive recycling endosome pathway along microtubules (Grimsey *et al.*, 2016), which functions to transport vRNPs to the plasma membrane to stimulate production of viral progeny (Stone *et al.*, 2016).

The above examples demonstrate YB1's role in promoting virus infectivity and production dynamics. Alternatively, YB1 has been reported to regulate viruses in a negative manner. Paranjape and Harris (2007) identified YB1 binding to Dengue Virus (DV) RNA, especially to the regulatory region of DV RNA 3'-UTR structure called the 3'-stem loop that plays a role in viral replication. Furthermore, Paranjape and Harris (2007) generated YB1^{+/+} and YB1^{-/-} mouse embryonic fibroblasts (MEF) from transgenic mice models, and infected these with DV strain 16681. YB1^{-/-} cells produced higher quantities of DV progeny compared to YB1^{+/+} cells. Further functional analyses indicated that YB1 played an antiviral role to DV infection in a complex manner involving, in part, binding of YB1 to 3'-UTR, repressing translation of viral proteins and/or repressing DV RNA replication.

Adenovirus relies on the expression of viral proteins encoded by *E1* and *E2* genes for replication. YB1's involvement in pre-mRNA splicing was first recorded by Chansky *et al.* (2001), and the splicing of pre-mRNA to mRNA, encoding E1A Adenovirus helper protein, was positively regulated by YB1 (Chansky *et al.*, 2001). The overexpression of YB1 in NIH3T3 cell lines promoted the alternative splice variants of E1A; namely influencing selection of E1A pre-mRNA splice sites for the 13S and 12S isoforms. Although, the above study did not look into the pre-mRNA splicing effect by YB1 in the context of Adenovirus infection, replication and/or production. However, a more direct involvement of YB1 with Adenovirus was demonstrated by studies using Adenovirus vectors and immunofluorescence by Holm *et al.* (2002). Adenoviral E1B was found to be co-targeted with YB1 to the nuclear compartment by Holm *et al.* (2002), and co-localised to nuclear bodies thought to be key sites for viral transcription and replication. Additionally, YB1 was shown to positively regulate gene expression, at least in part, from the late E2 promoter (Holm *et al.*, 2002). Overall, having reviewed YB1's role in the context of viral biology using specific examples, and outlining the relevance of YB1 in rAAV vector biology in section 1.3.7, YB1 seems to function in a virus specific-manner – YB1 is either hijacked by viruses which exploit YB1's DNA, RNA or protein binding functions to facilitate important steps in the virus life-cycle. Alternatively, YB1 seems to possess antiviral properties that limit certain virus' production and replication.

1.5.0 CRISPR and genome editing

1.5.1 CRISPR-Cas9 system – evidence of a prokaryotic adaptive immune system

The stress of dynamic, ever-changing microbiological communities has ensured the evolution of very effective coping mechanisms. These have ensured that dominant bacteria and archaea are able to host a number of natural habitats, even when these are often associated with challenging pressures. Pressures include naturally occurring biotic stresses that affect the survivability of a given bacterial/archaeal micro-organism. Biotic stresses, such as transmissible genetic elements (cosmids, plasmids) and predatory viruses, exert a constant pressure and pose a considerable threat. In addition, the exerted pressures are further complicated by the enhanced mutation rates that viruses easily succumb to (Lauring *et al.*, 2013). Micro-organisms, therefore, require co-evolving inherent defence mechanisms to respond to the variability of predatory viruses and transmissible genetic elements, to ensure their survival.

A regulatory system inherent in a number of bacterial and archaeal species termed the clustered regulatory interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein (Cas), is one such mechanism. The CRISPR-Cas system(s) describes an adaptive immune system that responds to the surveillance of foreign genetic elements that invade host micro-organisms (Labrie *et al.*, 2010). As a consequence, the process leads to successful immunisation against subsequent invasion by exogenic species. This is achieved by uptake of and loci-specific integration of foreign DNA into CRISPR loci. These loci comprise of short, alternating, interspaced, near-palindromic DNA repeats (varying between 23-55bp). Interspacing sequences (21-72bp) show hyper-variability. Given the near palindromic nature of these repeat sequences, post-transcriptionally these elements have great potential to form complex secondary structures in the form of RNA hairpins.

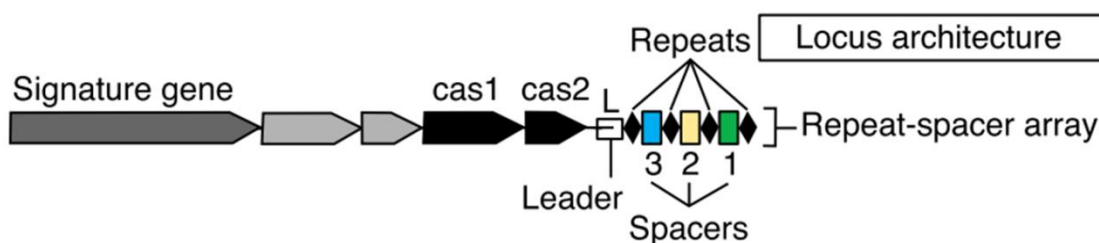


Figure 1.11 Generic architecture of a CRISPR locus. A CRISPR locus is composed of the CRISPR array that is made up of short repeat sequences (between 23-55bp in length), separated by unique spacer sequences (between 21-72bp in length). The CRISPR locus is further characterised by an AT-rich leader sequence (L) that precedes the CRISPR array. The CRISPR array is usually flanked by *cas* genes encoding Cas proteins – the functional effectors of CRISPR-Cas. The *cas1* and *cas2* genes are universally present within CRISPR systems, and signature *cas* genes define the type of CRISPR/Cas system – *cas3* for type I, *cas9* for type II, and *cas10* for type III. Figure from Barrangou (2015).

These loci were found to be accompanied by flanking *cas* genes (Fig. 1.11), which encode effector proteins that employ protective functions against invasive genetic elements (Jansen *et al.*, 2002). CRISPR loci were first identified in *Escherichia coli* K12 strain upstream to its alkaline phosphatase (*iap*) gene (Ishino *et al.*, 1987). Later more bacterial and archaeal species demonstrated frequent abundance of CRISPR loci – approximately half of bacterial genomes and approximately 83% of archaeal species (Grissa *et al.*, 2007). It is commonly recognised that most prokaryotic genomes possess a single locus, however, *Methanococcus* harbour multiple loci dispersed within its genomes (Grissa *et al.*, 2007). It has been more recently identified that CRISPR-like systems are also inherent to some larger giant viruses, which possess the genomic content to support CRISPR-Cas systems. A prototype example is mimivirus, of which lineage A type displays genomic content that resembles Zamilon (a unique virophage) as interspersed within an operon (Levasseur *et al.*, 2016). The presence of which locus correlated with resistance to the virophage (Levasseur *et al.*, 2016). Therefore, the associated system resembles a CRISPR-like system dubbed ‘mimivirus virophage resistance element’ (MIMIVIRE).

It was not until 2005 that the variable spacer sequences within CRISPR loci were identified as homologous to exogenous genetic sequences, originally derived from plasmids and viruses (Bolotin *et al.*, 2005; Pourcel *et al.*, 2005). Pourcel *et al.* (2005) dissected the spacer sequences of *Yersinia pestis* strains and determined that a number of spacer sequences were homologous to sequences derived from prophages (Pourcel *et al.*, 2005). Most instances revealed sequence homology specific to phage (Shmakov *et al.*, 2017). This mechanism of adaptation to foreign phage invaders was first observed in *S. thermophilus* (Barrangou *et al.*, 2007). Infection of *S. thermophilus* with phage resulted in the acquisition of phage-specific DNA in phage-resistant or surviving mutants. Consequently, the sensitivity of tested prokaryotic species to phage inversely correlates with the number of phage-homologous spacer elements (Bolotin *et al.*, 2005).

The ability to acquire, integrate, or expel exogenic material in a loci-specific manner provides a means to respond to highly mutable and evolving viruses (Jansen *et al.*, 2002; Makarova *et al.*, 2002). The system can be described as a highly dynamic immune system that maintains evolutionary sustenance against competing microorganisms. Similar is observed between eukaryotic organisms and the interactions these have with prokaryotic pathogens assumed under the Red Queen hypothesis (Van Valen, 1973).

The context in which CRISPR-Cas systems operate is governed by the upstream *cas* genes, which encode effector proteins necessary for CRISPR-Cas-mediated immunity. Originally it was suggested that these proteins functioned in a previously uncharacterised DNA repair system (Makarova *et al.*, 2002). Next, it was postulated that the CRISPR-Cas system functioned as a novel adaptive immune system inherent to prokaryotic micro-organisms, because of its resemblance to eukaryotic RNA interference (RNAi) system. The RNAi system promotes an additional level in eukaryotic gene regulation; it represents a post-transcriptional process, where the generation of dsRNA leads to gene silencing (Sijen *et al.*, 2001). Long dsRNA is recognised by Dicer, a nuclease of the RNase III family (Bernstein *et al.*, 2001; Doi *et al.*, 2003; Myers *et al.*, 2003). Dicer then processes the long dsRNA into smaller dsRNA molecules called siRNA (Zamore *et al.*, 2000; Elbashir *et al.*, 2001). These are able to bind to a number of RNase-activity containing proteins and form a complex called the RNA-induced silencing complex (RISC). RISC functions to help siRNA to target RNA in a sequence-specific manner for degradation (Hammond *et al.*, 2000; Martinez *et al.*, 2002). MicroRNAs (miRNAs) are another form of short non-coding (nc)RNA that also function post-transcriptionally to silence gene expression by inhibiting protein synthesis, similarly to that of siRNAs (Zeng *et al.*, 2003). Dicer-mediated maturation of miRNA allows it to bind to 3'-UTRs of mRNAs, followed by RISC recruitment for nuclease-mediated degradation (Chendrimada *et al.*, 2007). Some analogy between CRISPR-Cas and eukaryotic RNAi systems was inferred. Common features include the use of small ncRNAs that function in tandem with RNPs to facilitate sequence-specific cleavage. Despite the fact that CRISPR-Cas-mediated interference can target RNA, it is more commonly associated with targeting dsDNA (Jiang *et al.*, 2015).

1.5.2 Classifications of CRISPR systems in prokaryotes

Despite the genetic diversity of the *cas* genes (Jansen *et al.*, 2002; Makarova *et al.*, 2002), analyses of these genes suggested inclusive domains that are characteristic to endo- and exonucleases, helicases, RNA- and/or DNA- binding proteins (Jansen *et al.*, 2002; Makarova *et al.*, 2002). Originally, only four *cas* genes (*cas1-4*) were found in prokaryotes that harbour CRISPR loci. However, approximately 65 distinct Cas orthologs have been identified since, and work presented by Haft *et al.* (2005) has helped classify these into a collection of 45 families.

Generally, *cas1* and *cas2* genes are universally present within all CRISPR systems (Makarova *et al.*, 2011a). Identification of signature *cas* genes and genomic organisation has helped classify

CRISPR systems. In fact CRISPR systems have been classified into three main types (types I-III), which are distinguished by their unique *cas* gene content. Types I-III CRISPR-Cas systems are defined by the presence of *cas3*, *cas9*, and *cas10*, respectively (Makarova *et al.*, 2011b; Gasiunas *et al.*, 2012; Makarova *et al.*, 2015). Given the engineering and utilisation of type II CRISPR systems and Cas9 nuclease for genome editing, the type II system will be the principal focus for this review.

Type II CRISPR systems contain the repeat-spacer array as described above, but only contain three to four *cas* genes. *cas1*, *cas2* are required for the acquisition of spacer elements within the CRISPR locus. This is mediated, in part, by the required protospacer adjacent motif (PAM) sequences that are located downstream of the protospacers. PAM is required for target recognition, cleavage and then the acquisition of incoming DNA elements. The *cas9* gene predominantly classifies the CRISPR system as type II, and exact function will be detailed in a section 1.5.7; however, a unique feature that contributes to this CRISPR-type classification is the fact that Cas9 endonuclease is sufficient to mediate the effector functions of targeting and endonucleolytic cleaving foreign DNA, alone (Gasiunas *et al.*, 2012). This is unlike other CRISPR types I and III, which require recruitment of multiple proteins to form a functional Cas protein-CRISPR (cr)RNA complex – (cr)RNP (Brouns *et al.*, 2008). It is also noteworthy that the type II CRISPR system is further subtyped as type IIA, IIB and IIC, based on phylogenetic data (Makarova *et al.*, 2011a). Another distinguishing feature between the CRISPR types is also dictated by the biogenesis of crRNAs. The reliance on endogenous RNase III is required to process pre-crRNA transcripts into the separate crRNA functional units in type II systems (Deltcheva *et al.*, 2011).

In principle, all CRISPR-Cas system's main molecular mode of immunity is conserved – in that RNA guided nucleases elicit the effector functions for adaptation and interference towards foreign and invading genetic elements. Specifically, the crRNP complex assembled in type I systems is called the CRISPR-associated complex for antiviral defence (Cascade) (Brouns *et al.*, 2008). In type II systems, the entire interference is mediated by RNA-guided Cas9 protein – a singular Cas protein instead of a multiprotein composition as seen with Cascade (Wei *et al.*, 2015a). Type III systems are further standalone in their classification, and instead type III systems encode Csm or Cmr Cas proteins for subtypes III –A or –B, respectively (Rouillon *et al.*, 2013; Staals *et al.*, 2013).

1.5.3 PAM sequence dependency and self-non-self discrimination

The recognition of foreign material does show some CRISPR-type-dependent differences. Although, a constant feature of distinguishing self from non-self involves recognition of sequence motifs in the form of PAMs. PAM sequences are not only utilised for interference by CRISPR systems, but also for self and non-self discrimination for target acquisition. The unregulated acquisition of self instead of non-self could prove detrimental to the prokaryote, and would reflect a scenario analogous to autoimmunity observed in higher eukaryotes, except that self DNA is recognised by the host organism's CRISPR-Cas system.

'Self' acquisition has been defined in prokaryotic systems – Stern *et al.* (2010) dissected the underlying properties of self-acquisition into CRISPR loci and surmised that the acquisition of self into CRISPR loci occurred in 0.4% of total spacers examined. Furthermore, self-acquisition was not an evolutionary conserved phenomenon amongst the repertoire of CRISPR-encoding organisms tested, and was unlikely to contribute to a regulatory system of endogenous gene expression (Stern *et al.*, 2010). Given the partial or full inactivation of the CRISPR loci associated with self-acquisition, it was suggested that in such scenarios self-acquisition was considered deleterious to the organism. Since, additional studies have been performed and demonstrate the contrary – CRISPR-Cas systems is utilised as a form of endogenous gene regulation (Sampson and Weiss, 2013).

Further checkpoints are proposed in CRISPR-Cas containing organisms, which explain, in part, the preference for foreign DNA over self DNA. Work by Levy *et al.* (2015) has shown that in *E. coli* the selection of protospacers, and therefore their acquisition, was replication-dependent. Genomic loci which correspond to predicted stalled replication forks and DSBs in foreign DNA were preferred for selection. These breaks would be repaired by endogenous helicase/exonuclease RecBCD complex, which functions to degrade broken DNA ends. This degradation occurs until the RecBCD meets a Chi site. Chi sites are profoundly more rich and frequent in the chromosomal DNA, and far less evident in the invading, foreign DNA (Dabert *et al.*, 1992; Halpern *et al.*, 2007). Chi sites are short stretches of sequence that functions to change the enzymatic action of RecBCD complex from exonuclease to helicase. In turn, this promotes HR mediated by RecA instead as a DSB repair mechanism in prokaryotes (Dabert *et al.*, 1992). If no Chi sites are met, as is the case for RecBCD complexes in context with foreign DNA, it is thought that RecBCD generates candidate protospacer sequences that can then be processed further for spacer acquisition by its exonuclease function (Levy *et al.*, 2015).

However, the products of RecBCD exonuclease activity are variable ssDNA fragments and protospacer selection requires a short, 33bp dsDNA fragment (at least in the Type I CRISPR-Cas systems).

1.5.4 Mechanism of CRISPR-mediated adaptive immunity

A key defining feature of CRISPR-Cas systems is *cas* genes that are associated with a given loci. The proteins in which they encode are extremely diverse, and are the main effector molecules that carry out the core functions of the adaptive immune system for the given prokaryote. The CRISPR-Cas system carries out its function in three well-defined steps, and is summarised by: i) adaptation involving the acquisition of foreign genetic material and its integration into CRISPR loci in a PAM-dependent manner; ii) crRNA biogenesis, whereby the adapted CRISPR loci is capable of transcribing the repeat elements and spacer DNA, and its processing into functional guide (g)RNAs, and iii) targeting of foreign nucleic acids and its interference (Fig. 1.12).

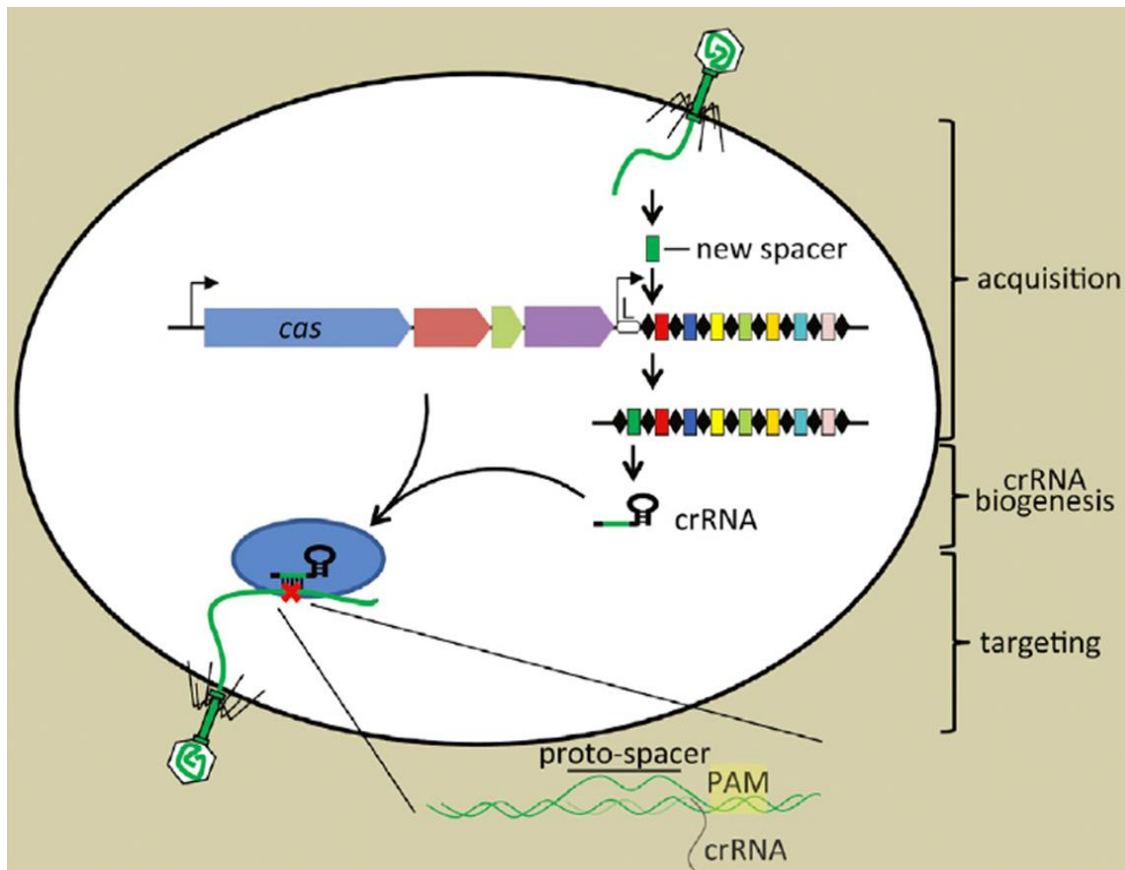


Figure 1.12 Three stages of CRISPR-mediated immunity in prokaryotes. CRISPR-mediated immunity in prokaryotes is instigated by acquisition of new spacer sequences from invading, foreign genetic material. Acquisition involves CRISPR-Cas-mediated integration of new spacer sequences into the CRISPR array, effectively immunising the CRISPR-containing organism against the invading species. Once integrated, the spacer sequence can be transcribed from the CRISPR array, with eventual processing results in crRNA biogenesis. The mature crRNA facilitates targeting of invading genetic elements, by binding to Cas effector protein(s), that recognises target sequence with suitable protospacer adjacent motifs (PAM) and that is complementary to crRNA. The end result is the degradation of the target DNA. Figure from Barrangou and Marraffini (2014).

1.5.5 Mechanism of spacer acquisition into CRISPR loci

Establishing a record of a previous infection of phage or plasmid by the CRISPR-Cas system is dictated by the adaptation and spacer acquisition machinery. This effectively immunises the CRISPR-containing organism to subsequent infections. Alternatively, in the context where a pre-existing immunity is held and subsequent infection involves a variant strain, then acquisition is described as priming (Datsenko *et al.*, 2012).

Spacer acquisition has been experimentally observed in various CRISPR types, and as a result the minimum prerequisites for the adaptation process have been identified. Some variations between CRISPR-Cas subtypes have been noted. The adaptation and spacer acquisition process can be generally described to involve a multistage procedure that first identifies foreign genetic material, followed by the insertion of DNA into the CRISPR loci as a new

spacer. The discrimination of acquisition of self from non-self is made possible by the presence of recognisable PAM sequences and dependency on DNA absent of Chi sites. Spacer acquisition generally occurs at the leader sequence end of the CRISPR locus with duplication of repeat sequence to maintain the repeat-spacer-repeat configuration, necessary for downstream crRNA biosynthesis (Wei *et al.*, 2015a).

Studies on spacer acquisition have been performed across the CRISPR types I-III, but the most comprehension on spacer acquisition has been demonstrated from studies involving type IE and type IIA systems. Recognition of new protospacer sequences is mediated predominantly by Cas1-Cas2 complex via interactions with PAM sequences. In fact, close conservation of the *cas1* and *cas2* genes is observed between CRISPR systems, indicating that a common mechanism is likely used. Currently it is understood that Cas1 of the complex is the main functioning component. Cas1 is a metal dependent nuclease-helicase (Wiedenheft *et al.*, 2009), and Cas2 seems to possess a structural role, instead to bridge and bind Cas1 proteins and act as a scaffold for the substrate DNA duplex (Wang *et al.*, 2015c). Cas2's metal-dependent nuclease activity showed no significant contribution to the adaptation phase. Mutations within Cas1's active site correlated with abolished integration of new spacers into CRISPR loci, whereas the catalytic activity of Cas2 was non-essential (Nunez *et al.*, 2014). When present, however, enhanced acquisition activity was observed (Nunez *et al.*, 2014).

E. coli's type IE CRISPR system involves Cas1 and Cas2 forming a stable heterohexameric complex (Wang *et al.*, 2015c), which is composed of two Cas1 dimers that are bound to either side of a Cas2 dimer. This heterohexameric complex has been shown to function in both acquisition of the protospacer and its integration into the CRISPR locus. Firstly, Cas1-Cas2 complex recognises the PAM sequence as 5'-AAG-3' and the PAM-complementary sequence (Wang *et al.*, 2015c). Substrate selection is further discriminated by the requirement for a 23bp duplex DNA with 3'-overhangs. Complete binding encourages a conformational change in the Cas1-Cas2 complex (Wang *et al.*, 2015c). Cas1-mediated cleavage of the protospacer's 3'-overhangs generates 5nt long overhangs with free 3'-hydroxyl groups. These are required for integration into the leader end of CRISPR locus (Nunez *et al.*, 2015). The proximal repeat is recognised by the Cas1-Cas2-protospacer complex and staggered cuts are proposed to be generated, the cuts of which are highly regulated by the presence of molecular rulers encoded in the repeat element (Goren *et al.*, 2016; Wang *et al.*, 2016b). Therefore, with the 3'-hydroxyl groups, nucleophilic attacks occur on the plus and minus strand of the repeat sequence at

either junction ends, creating two half-site intermediates. The flanking repeat sequences are DNA repaired by as yet uncharacterised DNA polymerase and ligase (Yosef *et al.*, 2012; Nunez *et al.*, 2015; Rollie *et al.*, 2015; Wang *et al.*, 2016b). The orientation of the protospacer acquisition process allows for the 3'-cytosine derived from the PAM-complementary sequence to attack the phosphate group of the distal repeat junction, to generate a spacer with a 5'-guanine as the first nucleotide in the complementary sequence (Nunez *et al.*, 2015).

Type I systems rely on protospacer acquisition and integration by Cas1-Cas2 complex. However, CRISPR-Cas type IIA systems rely additionally on *trans*-acting CRISPR (tracr)RNA, and Cas9, Cas1, Cas2, Csn2 Cas proteins. It is further noted that despite Cas1-Cas2 share conservation between CRISPR-Cas types, Cas9 in CRISPR-Cas type IIA predominantly performs the role of PAM recognition for protospacer selection (Heler *et al.*, 2015; Wei *et al.*, 2015b). The importance of this function is reflected by mutations in Cas9's PAM recognition domain. Mutations result in the acquisition of non-PAM containing spacer sequences (Heler *et al.*, 2015). Nonetheless, Cas9-dependency likely encourages the recruitment of the remainder Cas proteins to facilitate integration in a similar manner as type IE systems, thereafter (Wei *et al.*, 2015b).

1.5.6 Mechanism of crRNA biosynthesis

Completely mature crRNA is required for efficient targeting of foreign DNA. The transcription and processing of the CRISPR loci follows a general theme between CRISPR-Cas types – transcription is driven by an internal promoter located within the leader sequence (Pul *et al.*, 2010). This results in the generation of a long pre-crRNA transcript, which is processed by Cas endoribonucleasases to cleave the pre-crRNA at specific sites within the repeat regions (Li *et al.*, 2013; Wakefield *et al.*, 2015). This forms the required downstream gRNAs for crRNP formation and target DNA recognition. Similarities in pre-crRNA processing have been observed between CRISPR-Cas types, especially between types I and III. For instance, almost all type I and III CRISPR systems utilise Cas6 protein (or a closely related orthologue) for pre-crRNA processing (Li *et al.*, 2013; Wakefield *et al.*, 2015). Therefore crRNAs that are formed are composed of the central spacer RNA sequence which is complementary to the target DNA sequence flanked by some repeat sequences (Carte *et al.*, 2008; Li *et al.*, 2013).

Type II processing of crRNA shows some differences from that observed in type I and III systems, where the coordinated action of the *trans*-acting tracrRNA, endogenous RNase III and

Cas9 protein function together to process pre-crRNA into functional gRNAs. Specifically, the tracrRNA was identified by differential RNA sequencing by Deltcheva *et al.* (2011) on a type II CRISPR-Cas locus in *Streptococcus pyogenes*. The small RNA is encoded upstream to *cas* genes on the opposite strand within the leader sequence of type II CRISPR loci (Deltcheva *et al.*, 2011). Most strikingly, this study identified primary long tracrRNA species, which showed some sequence complementarity to the repeat sequences present within pre-crRNA sequence, potentiating base-pairing with pre-crRNA transcripts. Differential RNA sequencing of co-processed tracrRNA and pre-crRNA also exhibited short 3'-overhangs, which were indicative of endoribonuclease cleavage by RNase III (Deltcheva *et al.*, 2011). In fact, crRNA biogenesis was highly dependent on the activity of endogenous RNase III, as mutations in the RNase III-encoding *rnc* gene or in the repeat complementary sequences resulted in the inability to process tracrRNA and pre-crRNA into intermediate or mature species.

Following stabilisation of tracrRNA:pre-crRNA interactions by Cas9 (Karvelis *et al.*, 2013), RNase III is recruited and cleaves within the repeat sequences to form individual crRNA units, (Deltcheva *et al.*, 2011). The 5'-end of the crRNA unit is further trimmed down (Deltcheva *et al.*, 2011), with the loss of some spacer sequence. The remainder of the spacer sequence is complementary to the target DNA, and is sufficient to successfully target invader DNA for degradation in a Cas9-dependent manner. Type IIC systems, such as that seen in *N. meningitidis* (Zhang *et al.*, 2013), the individual crRNA units are instead generated separately due to the presence of separate promoters within or around the repeat sequences that drives transcription of these single pre-crRNA units. Though RNase III-mediated processing can occur for these individual pre-crRNA units, RNase III was shown to be relatively unnecessary. RNase III-deficient *E. coli* (*rnc*-) was capable of processing of CRISPR array that included only a single spacer (Karvelis *et al.*, 2013). Type IIC CRISPR-Cas from *N. meningitidis* possesses separate promoters between each spacer for individual crRNA expression, and does not require additional processing (Zhang *et al.*, 2013). In these instances, tracrRNA and RNase III was required for 3'-end formation, but RNase III was considered dispensable in this type IIC CRISPR-Cas system (Zhang *et al.*, 2013). This perhaps illustrates an interesting form of evolution of the CRISPR-Cas system, where type IIC-containing bacteria have evolved means to target specific phage infections that would otherwise inhibit endogenous RNase activity. In fact, the giant phage ϕ KZ encodes RNase inhibitor gp37/Dip, which has been shown to inhibit RNA degradosome formation in *Pseudomonas aeruginosa* (Van den Bossche *et al.*, 2016).

1.5.7 Mechanism of CRISPR interference

Therefore, with the biosynthesis of functional guide crRNAs specific to target invading DNA molecules, the CRISPR systems are capable of providing robust and adaptive immunity in the form of interference. The process of interference involves the crRNA elements to guide the recognition and precise cutting of target DNA by effector Cas nucleases. This is made possible by complex of the Cas nuclease(s) to crRNA to form a crRNP molecule. The crRNA-‘programmed’ Cas nuclease is specifically guided to and recognises target DNA by way of PAM sequences, which in turn instructs Cas proteins to perform their function to degrade target DNA.

Type I interference has been modelled principally by *E. coli* type IE system, and the mechanism of interference involves the formation of the crRNA-guided Cascade complex to perform foreign DNA surveillance and interference (Jore *et al.*, 2011). Cascade assembly is initially comprised of five different Cas proteins in various subunit stoichiometry: (Cse1)₁, (Cse2)₂, (Cas5)₁, (Cas7)₆, and (Cas6)₁ (Brouns *et al.*, 2008; Jore *et al.*, 2011). This complex of Cas proteins assemble upon the 61nt long crRNA. Final structural analysis gave the Cascade crRNP a structure that was divided into parts – the head, belly and tail, with an overall resemblance to a seahorse. Maturation of the 61nt long crRNA, by Cas6-mediated endoribonuclease cleavage within the repeat sequence of the pre-crRNA transcript (Li *et al.*, 2013; Wakefield *et al.*, 2015), generates a final crRNA product which is composed of a central 32nt spacer sequence, a 21nt long 3'-hairpin structure and 8nt 5'-end (Jore *et al.*, 2011). Cas6 remains bound to the 3'-hairpin structure of the mature crRNA (Jackson *et al.*, 2014), forming the head of the assembled structure. This is followed by the rest of Cascade assembly by recruitment of Cas5 to the 5'-end and association with the large subunit (Cse1) to form the tail of the Cascade complex (Jackson *et al.*, 2014). The six units of Cas7 bind to the spacer sequence and form a helical backbone, effectively bridging the head and tail (Jore *et al.*, 2011). The small Cascade subunit (a dimer of Cse2) is also recruited and forms the belly, and functions to stabilise both target DNA and crRNA (Jackson *et al.*, 2014).

In order to carry the final step of interference, Cascade first recognises the PAM sequence that is present in target DNA. This is mediated by interactions between the target DNA and Cse1 (Tay *et al.*, 2015; Xue *et al.*, 2017). Target duplex is then unwound and crRNA binds to complementary, target DNA sequence (van Erp *et al.*, 2015). Conformational change of Cascade and the formation of a DNA R-loop in the target, foreign DNA ensue, which permits

the recruitment of Cas3 helicase-endonuclease (Mulepati *et al.*, 2014). Cas3 is the main effector enzyme that nicks foreign ssDNA, resulting in a 200-300nt gap adjacent to the crRNA-complementary sequence (Redding *et al.*, 2015). Cas3 continues to translocate the bound DNA in the 3'>5' direction to processively unwind and degrade it (Gong *et al.*, 2014; Redding *et al.*, 2015).

Unlike the type I system which requires a multi-subunit complex to elicit its effector functions, type II systems requires only a single protein (Cas9) to function in cohort with crRNA (Gasiunas *et al.*, 2012). However the addition of tracrRNA bound to Cas9 is essential to perform target DNA recognition and degradation (Deltcheva *et al.*, 2011; Chylinski *et al.*, 2013). Crystal structures of type II crRNP belonging to *S. pyogenes* has been examined (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014), and provided insight of overall mechanism of action. *S. pyogenes* (Sp)Cas9 was found to be approximately 100-190kDa, with ultrastructural analysis revealing a bi-lobed protein composed of a variable α -helical lobe for binding to nucleic acids, a nuclease lobe that harbours the catalytic RuvC and HNH nuclease domains, and a C-terminal topoisomerase homology domain (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014). In order to target foreign DNA, Cas9 nucleic acid binding region associates with the gRNA (dual crRNA:tracrRNA). Cas9 binds to the guide portion of the crRNA in a sequence-independent manner, and also to the duplex region of the gRNA in a sequence-specific manner (Jinek *et al.*, 2012). Additionally, Cas9 binding with gRNA or single-gRNA (sgRNA) is required to activate the scanning capabilities of Cas9 for PAM recognition (Jinek *et al.*, 2012; Jinek *et al.*, 2014). Local melting of target DNA ensues and complementary base-pairing occurs between crRNA and target DNA (Anders *et al.*, 2014; Mekler *et al.*, 2017). Target DNA is cleaved on both strands by the nuclease lobe of Cas9 – HNH nuclease domain cleaves the strand complementary to the crRNA, while the RuvC nuclease domain cleaves the non-complementary DNA strand, generating a DSB, precisely 3nt upstream from the PAM sequence (Fig. 1.13) (Gasiunas *et al.*, 2012).

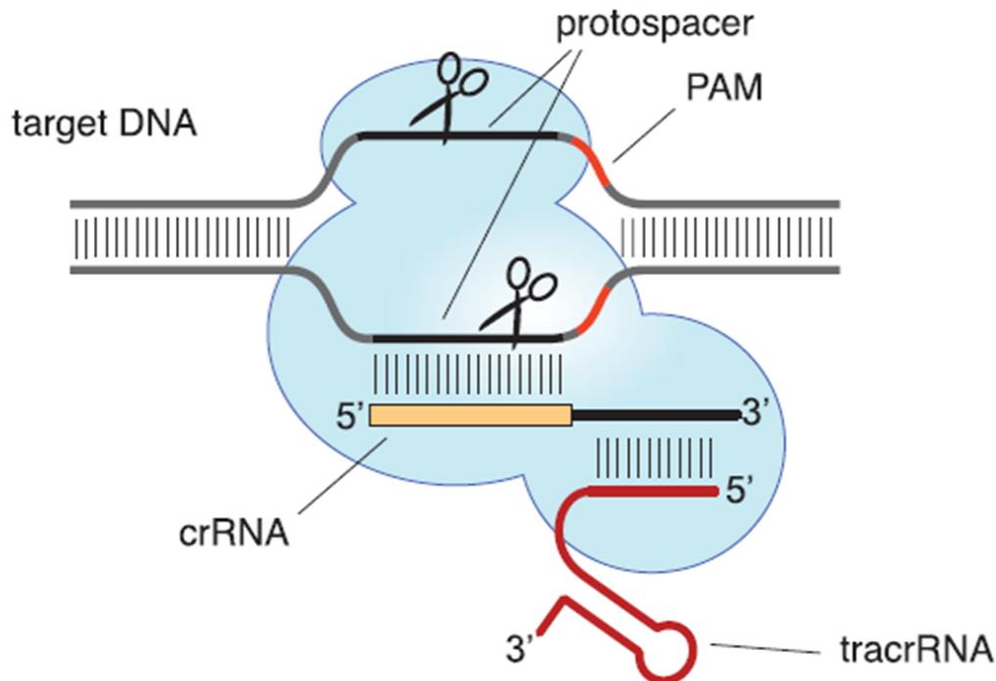


Figure 1.13 Cas9 mechanism of CRISPR interference of target DNA. The dual-RNA (gRNA) is composed of the crRNA and tracrRNA, and in its mature state is able to complex with Cas9 endonuclease. Cas9 is able to scan target DNA for the PAM, after which local melting of the DNA duplex occurs. This allows the crRNA sequence to bind to protospacer sequence by complementary base-pairing. Cas9 is then able to cut the target DNA and cause a DSB, by nicking the complementary and non-complementary DNA strands, 3nt upstream of the PAM. Figure from Jinek *et al.* (2012).

1.6.0 From adaptive immunity to targeted genome editing *in vitro* and *in vivo*

1.6.1 Definition of targeted genome editing

Genome editing describes a process involving the targeting of genomic DNA sequence, and incorporating a desired change(s). The use of targeted gene modifications has a number of advantages in biomedical, biotechnology, and basic biological research. Techniques and technologies that allow for precise and targeted deletions, insertions or modifications can allow complete characterisation of specific genes and their function.

Gene targeting was perhaps initially explored by research introducing transgenes into the genome in a targeted manner. This required HR to facilitate such process (Folger *et al.*, 1982); however, the main facilitator of gene targeting was found to be DSBs at/or around the desired loci. This helped drive endogenous HR-mediated DNA incorporation when a donor DNA template was present. The importance of induced DSBs was first impacted by model experiments done in yeast (Rudin *et al.*, 1989), and mammalian cell lines (Rouet *et al.*, 1994; Choulika *et al.*, 1995). These studies showed that inducing DSBs with highly specific

endonucleases facilitated the repair of the DSB by recombinogenic mechanisms. These induced DSBs instigated the gene targeting of an exogenous, homologous gene sequence to function in the repair of the DSBs. Desired genome alterations was made possible in a targeted fashion, albeit inefficiently. The frequency of the desired recombination was very poor – 1 in 10^6 - 10^9 cells (Capecchi, 1989). Alternatively, when an exogenous, homologous gene sequence was absent, then these induced DSBs would instead incorporate insertions and/or deletions (indels) due to the non-homologous end joining (NHEJ) pathway of DNA repair (Bibikova *et al.*, 2002).

1.6.2 Features of type II CRISPR-Cas system utilised for genome editing

In more recent years, programmable nuclease-based technologies were established as the more prominent and efficient methodology for targeted genome editing. The CRISPR/Cas9 system, in particular, has been most rapidly developing and applied for manipulation of mammalian and non-mammalian genomes.

As a genome editing tool, CRISPR-Cas9 type II system has been mostly adopted given the Cas9 nuclease ability to function alone with no need for other Cas proteins (Gasiunas *et al.*, 2012). Cas3 from the type I system, in contrast to Cas9, degrades target DNA processively. This was less useful for precise genome editing, especially when processive degradation of the target DNA could not be fully controlled for the intention of exact genome editing. Additionally, the type I system requires a machinery of different Cas proteins as a minimum requirement to target a specific DNA sequences. On the other hand, Cas9 generates a single DSB at the target sequence, which is an important requirement of genome editing systems to date. This results in a simple and programmable genome targeting system for desired genome editing.

Work performed by Jinek *et al.* (2012) on the type II CRISPR-Cas system identified that Cas9 recognised target foreign DNA for cleavage via a combined dual-RNA structure composed of the crRNA and tracrRNA. The crRNA contains a sequence that is complementary for a portion of the tracrRNA, forming a region of duplex (Jinek *et al.*, 2012). The crRNA alone in the presence of purified Cas9 was unable to cleave plasmid DNA harbouring the crRNA-complementary sequence with the necessary PAM. When tracrRNA was also present Cas9 was active and capable of cleaving the plasmid DNA at the intended site. This led to the understanding that the crRNA:tracrRNA dual-RNA structure was necessary for Cas9-mediated targeting. Additionally, the minimum tracrRNA sequence was dissected, which still encourages

Cas9 functionality, and in turn paved way for the development and utilisation into a programmable genome editing tool (Deltcheva *et al.*, 2011; Jinek *et al.*, 2012). Jinek and colleagues designed chimeric RNA that encoded both the 5'-crRNA and 3'-tracrRNA (Jinek *et al.*, 2012). The chimeric RNA (otherwise known as sgRNA) was permissible for Cas9-mediated DNA cleavage of target DNA that contained the crRNA complementary target sequence. These redefined sgRNAs further simplified the use of the CRISPR/Cas9 system for DNA cleavage and genome editing.

Another attractive feature of using SpCas9 for targeted genome editing is the simple PAM requirement of 5'-NGG-3' adjacent to the target sequence on the 3' end (Anders *et al.*, 2014). This is in contrast to other Cas9 homologues such as those derived from *Staphylococcus aureus* (SaCas9), *Streptococcus thermophilus* (St -1 or -3 Cas9), which have PAM requirements of 5'-NNGRRT-3' (Nishimasu *et al.*, 2015), 5'-NNAGAAW-3', and 5'-NGGNG-3' (Karvelis *et al.*, 2015), respectively. The shorter PAM sequence requirement demonstrated by SpCas9 significantly increases the number of potential genomic targets, as longer PAM sequences with relatively strict sequence motif limitations are less likely to occur naturally in target genomes.

Therefore, altogether, the CRISPR/Cas9 system for genome editing requires only the Cas9 endonuclease, the sgRNA with specificity to the intended target site, and/or a donor template. The CRISPR/Cas9 system operates similar to the aforementioned interference mechanism that is native for type II CRISPR systems - a DSB is generated precisely 3nt upstream of the 5'-NGG-3' PAM sequence (Fig. 1.14A). Principally, it is the host cells' own DSB repair mechanism that performs the genome editing – NHEJ promotes the formation of indels, but HR is required if a desired mutation (site directed mutagenesis) or DNA sequence (knock-in) is to be incorporated (Fig. 1.14B).

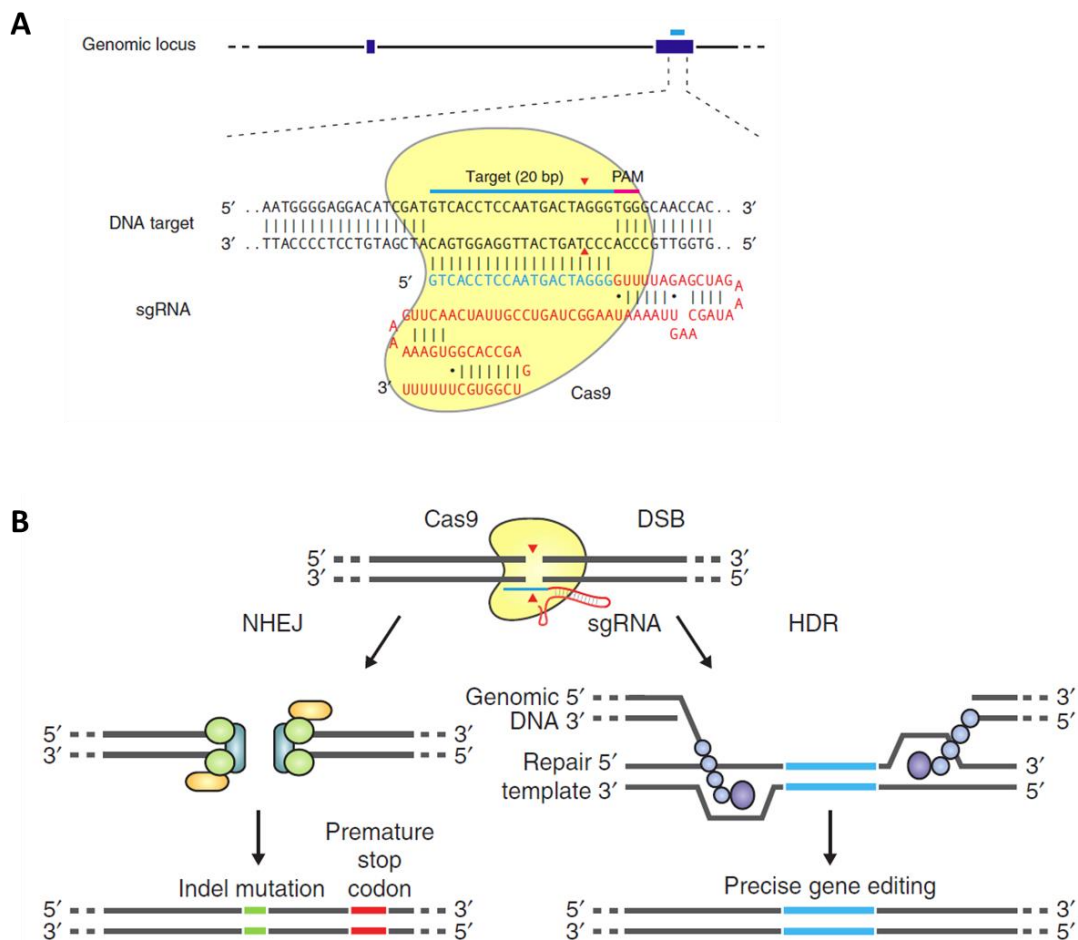


Figure 1.14 CRISPR/Cas9 genome editing of target DNA. A) Genomic loci can be targeted using CRISPR/Cas9 system, involving the association of sgRNA with Cas9. Cas9 is guided to target genomic locus due to presence of PAM and complementary binding between crRNA sequence and DNA target sequence. As a result, Cas9 is able to generate a DSB precisely 3nt upstream of the PAM sequence (see red arrows). **B)** After Cas9 generates a DSB, the host cell DNA repair machineries perform genome editing. NHEJ occurs when no donor DNA template is present. NHEJ repairs DSBs in an error-prone manner, leading to generation of indels and frameshift mutations. HR occurs when donor DNA template is present, and recombination between the DSB and donor DNA template allows the incorporation of desired mutations, for precise genome editing. Figures from Ran *et al.* (2013b).

1.6.3 The development of CRISPR/Cas9 system for genome editing and gene manipulation

The CRISPR/Cas system for genome editing has developed considerably since the characterisation of CRISPR systems in prokaryotes (Ishino *et al.*, 1987), and the demonstration that the CRISPR/Cas9 system can be engineered to target specific sequences in eukaryotic cell lines (Cong *et al.*, 2013; Mali *et al.*, 2013b). The initial adaptation of CRISPR/Cas9 technology for genome editing involved cloning codon optimised SpCas9 (with NLS) and the chimeric crRNA and tracrRNA sequences into plasmids for transfection into mammalian cell lines. Since then, a public depository called Addgene (<https://www.addgene.org/crispr/>), and now biotechnology development companies, including, but not limited to, Thermo Fisher Scientific

(<https://www.thermofisher.com/uk/en/home/life-science/genome-editing/geneart-crispr.html>) and Sigma-Aldrich (<https://www.sigmaaldrich.com/catalog/product/sigma/crispr?lang=en®ion=GB>) have made available CRISPR/Cas9 vectors, to provide the genome editing tool to virtually all research institutions.

A large concern of CRISPR/Cas9 genome editing was off-target activity by Cas9 – in that sgRNA designs used to edit a desired gene could redirect Cas9 to an off-target sequence. Off-target DSBs could be followed by unintended mutations. Experimentally, the fidelity of Cas9 was analysed and minimum off-target sites were identified for the given targeted genomic sequences in mammalian cell lines (Kim *et al.*, 2015; O'Geen *et al.*, 2015; Tsai *et al.*, 2015). This is due to the flexibility Cas9 has for gRNA:DNA complementarity; as few as three mismatches can be tolerated (Mali *et al.*, 2013a). Therefore, the propensity for off-target influences is possible when using Cas9, meaning improving the genome editing system was essential. Initial attempts to improve its fidelity included Cas9 modifications, namely the D10A or H840A substitutions in the RuvC and HNH nuclease domains, respectively, creating Cas9 nickase (Cas9n) variants (Ran *et al.*, 2013a). These Cas9 mutants functioned by generating a single nt nick only on the PAM-containing or PAM-complementary DNA target sequence, respectively. Therefore, a second sgRNA that was specific to another PAM sequence on the opposite strand was required to generate a staggered DSB (Ran *et al.*, 2013a). In theory Cas9n would reduce off-target effects, because off-target sites were unlikely to share complementary DNA sequences for both sgRNAs in close proximity and on opposite strands to each other. Additionally, single nt nicks are easily repaired with high fidelity by base-excision repair compared to DSBs and NHEJ (Shen *et al.*, 2014). When Cas9 possessed both the D10A and H840A mutations, this generated a nuclease-deficient, dead Cas9 (dCas9), which retained the DNA targeting function. Improved targeting efficiencies have also been observed when dCas9 was fused to FokI nuclease at its N-terminus (Guilinger *et al.*, 2014; Wyvekens *et al.*, 2015). FokI nuclease requires dimerization in order to become catalytically active and cleave DNA. Therefore, a pair of FokI-dCas9:sgRNA is required to complex at the target sequence in order for FokI nuclease activation and subsequent mutagenesis.

Additional developments included modifications in the sgRNA, either in the tracrRNA-scaffold or target complementary sequences. Truncating the 5'-end of complementary sequence of the gRNA from 20nt to 17/18nt demonstrated reduced or undetectable off-target mutagenesis (Fu

et al., 2014). Extending the tracrRNA sequence to generate a longer crRNA:tracrRNA duplex region or utilising full length tracrRNA sequence, generated optimised sgRNA structures, and correlated with enhanced on-target activity and specificity (Dang *et al.*, 2015; Xu *et al.*, 2017).

A large research focus has also been invested in development of Cas9 itself other than the aforementioned Cas9n and dCas9 variants. This includes expanding the delivery mechanism of Cas9 nuclease with gRNAs. For instance, CRISPR/Cas9 genome editing is not limited to transfection or electroporation of plasmids into target cells, but delivery vehicles also include lentiviral (Doench *et al.*, 2014; Sanjana *et al.*, 2014) or AAV (Ran *et al.*, 2015) vectors. In the case of AAV vectors, the limited packaging capacity restricts the CRISPR/Cas9 system to SaCas9 (approximately 3.2kbp) (Ran *et al.*, 2015). Although, SpCas9 (approximately 4.2kbp) can technically fit, but this leaves very little room for sgRNAs and control elements to be encoded. Additionally, recombinant Cas9 protein is available to transfect or electroporate into target cell lines with either synthetic or *in vitro* transcribed sgRNAs (Liang *et al.*, 2015). The purpose of which is to create a genome editing platform based on CRISPR/Cas9 that is significantly more transient than the transfection of plasmids, which otherwise runs the risk of integrating into the genome.

SpCas9 requires the presence of the 5'-NGG-3' PAM sequence on the 3' of the target sequence. Although a simple PAM sequence, and likely found frequently throughout the target genome, SpCas9 remains 5'-NGG-3'-restricted. Sequences that do not possess 5'-NGG-3' PAM sequences are thus restricted, and include AT-rich regions. Granted there are additional Cas9 orthologues: SaCas9 or StCas9s with different PAM requirements. These help increase the range of target sequences for genome editing. However, their PAM requirements are stricter, and can still limit potential targets for genome editing. Given, also, that SpCas9 is the most commonly used Cas9 for research and genome editing, keeping to a single Cas9 orthologue for all CRISPR/Cas9 research and translational needs would enable far easier standardisation of the genome editing tool. The PAM sequence is recognised by a PAM interaction domain found in the C-terminal topoisomerase homology domain of Cas9 (Jinek *et al.*, 2014; Nishimasu *et al.*, 2014). Two arginine residues (R1333 and R1335) facilitate the interaction with atoms of the guanine nucleobases that make up the 5'-NGG-3' PAM (Anders *et al.*, 2014). However, simply introducing substitution mutations in SpCas9 gene sequence (corresponding to R1333Q and R1335Q) did not generate functional SpCas9 nucleases that cleaved at target sites with 5'-NAA-3' PAM (Anders *et al.*, 2014). Alternatively, engineered SpCas9 was generated by directed

evolution (Kleinstiver *et al.*, 2015). This resulted in the identification of select aa substitutions: D1135V, R1335Q, T1337R or D1135E, R1335Q, T1137R, and were required to generate two novel SpCas9 variants that were capable of recognising the altered 5'-NGA-3' PAM sequence *in vitro* (Kleinstiver *et al.*, 2015).

Alternatively, manipulation of gene expression has also been researched using modified Cas9 nucleases. Initially, dCas9 was found capable of repressing gene expression when targeted to gene promoter sequences (Qi *et al.*, 2013). This CRISPR-mediated interference (CRISPRi) was further adapted with dCas9-fusion proteins; for example dCas9 was fused with transcriptional repressors and chromatin modifiers, such as Krüppel associated box (KRAB), which resulted in substantial gene silencing of targeted endogenous genes when compared to dCas9 alone (Gilbert *et al.*, 2013). In contrast, upregulation of endogenous genes resulted from fusion of dCas9 with transcriptional activators such as VP64, creating the CRISPR-mediated activation (CRISPRa) system. However, earlier studies showed that multiple sgRNAs are encouraged per target to see markedly enhanced upregulation of target gene expression (Maeder *et al.*, 2013; Perez-Pinera *et al.*, 2013).

SpCas9 has been further engineered to increase its specificity for on-target sequences and reduce non-specific interactions at off-target sequences (Kleinstiver *et al.*, 2016; Slaymaker *et al.*, 2016; Lee *et al.*, 2017a). This has been achieved by close analysis of SpCas9 crystal structure in complex with gRNA and target DNA. Analysis suggested a positively-charged groove made up of 32aa, in which SpCas9 is thought to stabilise the PAM-complementary sequence (Slaymaker *et al.*, 2016). Select substitution mutations of this positively-charged groove with neutral alanine residues, generated enhanced specificity SpCas9 (eSpCas9) variants. As a result, the eSpCas9 variants require only the minimum binding energy between gRNA and target DNA to prevent re-hybridisation of target DNA strands. As a result, eSpCas9 showed improved on-target specificity due to undetectable off-target indels in the known off target sites against *EMXI* and *VEGFA* gRNAs in 293 (Slaymaker *et al.*, 2016). Alternative to this was alanine substitutions in key residues that make contact with the PAM-containing DNA strand instead. This generated SpCas9 high fidelity (SpCas9-HF) variants (Kleinstiver *et al.*, 2016), which demonstrated reduction in nearly all genome-wide off-target sites for gRNAs against *EMXI*, *FANCF*, *RUNXI* and *ZSCAN2* genes (Kleinstiver *et al.*, 2016). More recently, directed evolution of SpCas9 in *E. coli* has been used to identify SpCas9 variants that discriminate against off-targets sites and show preferential binding to on target sites (Lee *et*

al., 2017a). Using positive selection pressure of on-target cleavage of the plasmid encoded *ccdB* lethal gene (harbouring the on-target sequence), and negative selection pressure of off-target cleavage of *E. coli* chromosomal DNA (harbouring the off-target sequence), Cas9 variant (Sniper-1) was identified and showed improved specificity, even when compared against eSpCas9 and SpCas9-HF (Lee *et al.*, 2017a). However, this demonstrates a potential need for directed evolution of SpCas9 against any given gene target in competition against a plethora of possible off-target sequences *in vitro* to merit actual development.

Both type I and III CRISPR/Cas systems have been demonstrated as capable of performing genome editing. However, repurposing of these CRISPR/Cas systems for genome editing has mainly been explored in archaeon species and not mammalian cell lines (Li *et al.*, 2016; Cheng *et al.*, 2017a; Liu *et al.*, 2017). A Cas endonuclease (Cpf1) derived from the relatively new type V CRISPR-Cas system was identified (Zetsche *et al.*, 2015), and shows unique features that promoted its potential for programmable genome editing. First, DNA targeting with Cpf1 only required crRNA, unlike Cas9 which requires a dual-RNA composed of crRNA and tracrRNA. Second, Cpf1 recognises a T-rich PAM at the 5'-end of the target protospacer, namely 5'-TTTV-3' (Kim *et al.*, 2017), unlike SpCas9 which recognises 5'-NGG-3'. Third, Cpf1 generates a DSB with cohesive ends, whereas Cas9 generates blunt ends. Lastly, Cpf1 also possesses RNase III catalytic activity, required for endogenous pre-crRNA processing (Fonfara *et al.*, 2016), which is advantageous in multiplex genome editing from a single CRISPR array. Type II CRISPR-Cas9 system requires endogenous RNase III to process pre-crRNAs - a function that is not inherent to Cas9. Therefore, mature sgRNAs are encoded in CRISPR/Cas9 systems to off-set this limitation, but this does prevent multiplex genome editing from a single CRISPR array *in vitro*. Cpf1 orthologues from *Lachnospiraceae bacterium* ND2006 (LbCpf1) and *Acidaminococcus sp. BV3L6* (AsCpf1) have shown promising and demonstrable potential for genome editing in human and mammalian cell lines (Zetsche *et al.*, 2015; Kleinstiver *et al.*, 2016b; Toth *et al.*, 2016). With the unique features of type V CRISPR-Cas systems, Cpf1 further expands the genome editing and manipulation potential where SpCas9 fails or is restricted.

1.6.4 Development of CRISPR/Cas9 system for gene therapy and as a future novel medicine

Regardless of certain shortcomings reviewed above, the development of the CRISPR/Cas9 system has delivered a potent genome editing tool with remarkable applicability *in vivo*. We have mainly stressed the application of CRISPR/Cas9 and derivatives or variants thereof *in*

vitro; however, CRISPR/Cas9 genome editing is quickly being incorporated *ex vivo*, *in vivo*, and clinical research as a potentially novel class of medicine.

With high-throughput sequencing technologies, the relationship between genes and a number of human diseases has become clearer. Moreover, animal disease and cancer models are well established, and have permitted *in vivo* testing models to examine the viability of the CRISPR/Cas9 genome editing tool as a gene therapy. Interestingly, the CRISPR/Cas9 system has become a more reliable and robust tool to generate transgenic animal models, and outclassing conventional methodologies (Wang *et al.*, 2013; Huang *et al.*, 2017). Initial evidence that suggested that the CRISPR/Cas9 system could be exploited as a gene therapy was perhaps demonstrated with studies targeting HIV1 *in vitro*. CRISPR/Cas9 showed encouraging inhibition of the HIV1 expression machinery in HIV1 provirus-integrated human cell lines, by targeting the long terminal repeat (*LTR*) promoter sequences, specifically the *TAR* and *U3* regions (Ebina *et al.*, 2013). This was found to be, in part, due to the power of the CRISPR/Cas9 to excise the latent form of HIV1 pro-viral DNA from the host genome (Ebina *et al.*, 2013; Hu *et al.*, 2014) (Kaminski *et al.*, 2016), and significantly impacted on the replicative and reactivating potential of HIV1 infection *in vitro* (Liao *et al.*, 2015; Zhu *et al.*, 2015). Furthermore, Liao *et al.* (2015) conceptualised that stable HIV1-targeting CRISPR/Cas9 conferred significant resistance to subsequent HIV1 infection, especially in more physiologically and clinically relevant human primary CD4⁺ T cells and haematopoietic lineages, which function as latent HIV1 reservoirs.

Careful consideration in CRISPR/Cas9 target has been recently warranted – the emergence of HIV viral escape mutants has been reported after targeting using CRISPR/Cas9 (Wang *et al.*, 2016a; Wang *et al.*, 2016d; Lebbink *et al.*, 2017). In such cases, indels introduced by the CRISPR/Cas9 system were either deleterious to HIV1 or the virus emerged as mutated escape strains, which were refractory to CRISPR/Cas9 and the gRNA originally used (Wang *et al.*, 2016a; Wang *et al.*, 2016d; Lebbink *et al.*, 2017). This revelation warrants caution when targeting HIV1's latent pro-viral DNA *in vivo*. Although, a study performed by Kaminski *et al.* (2016) were incapable of identifying escape mutants *in vivo* in transgenic mice and rat models that carry a transgene derived from a replication-deficient HIV1 genome after targeting using CRISPR/Cas9.

An alternative strategy explored genome editing of induced pluripotent stem cells (iPSCs) using CRISPR/Cas9 and piggyBac technology to generate homozygous *CCR5*Δ32 mutation in

the C-C chemokine receptor type 5 gene (*CCR5*) (Ye *et al.*, 2014). The *CCR5* protein is utilised by HIV1 as a co-receptor to infect T-cells, macrophages and monocytes (Choe *et al.*, 1996). However, *CCR5* Δ 32 mutation was originally found in individuals who showed resistance to primary macrophage-tropic HIV1 (Liu *et al.*, 1996). Therefore, *CCR5* potentiated as a promising target for conferring HIV1 resistance *in vivo* and prevent the rebound of HIV load after discontinuation of anti-retroviral therapy (Hutter *et al.*, 2009). A curative, allogenic stem-cell transplant strategy was devised, albeit, represented by only a single treated individual. Introducing homozygous *CCR5* Δ 32 mutation in the *CCR5* gene using CRISPR/Cas9 resulted in HIV1 resistant cell lines, such that HIV1 p24 glycoprotein concentrations were undetectable in the genome edited iPSCs, post-differentiation into macrophages and monocytes (Ye *et al.*, 2014; Kang *et al.*, 2015). Perhaps the only caveat of this approach was the relatively low efficiency of generating and identifying iPSCs with homozygous *CCR5* Δ 32 mutation (33-41%); although, CRISPR/Cas9 performed significantly better compared to alternative nucleases such as TALENs (14%) (Ye *et al.*, 2014; Kang *et al.*, 2015). In turn, an autologous, *ex vivo* gene editing pipeline to generate differentiated cells that are resistant to HIV1 infection provides a promising strategy to combat HIV1 infection into remission, analogous to observations by Hutter *et al.* (2009). The antiviral therapeutic capacity of CRISPR/Cas9 is becoming more relevant for targeting other clinically relevant viruses. This includes, but not limited to: Hepatitis B Virus (Lin *et al.*, 2014; Seeger and Sohn, 2014; Li *et al.*, 2017), Epstein Barr Virus (Wang and Quake, 2014; Yuen *et al.*, 2015; van Diemen *et al.*, 2016), and HSV (van Diemen *et al.*, 2016).

Exploring CRISPR/Cas9's antiviral therapeutic potential has been limited mainly to *in vitro* and even fewer *in vivo* studies. The adaptation of CRISPR/Cas9 as a novel class of antiviral medicine requires much further exploration and optimisation before human clinical trials can be considered. That being said, early phase human clinical trials have been completed or are currently active, which are exploring the capacity for zinc-finger nucleases to vaccinate against HIV1 (Table 1.1). These studies perpetuate the potential for human clinical trials that could explore the feasibility of CRISPR/Cas9 genome editing as a novel, anti-viral strategy. Of particular note is clinical trial research data presented by Tebas *et al.* (2014), whose team engrafted twelve patients with autologous CD4⁺ T-cells that had been edited at *CCR5* using zinc-finger nucleases *ex vivo* to confer disruption of *CCR5* expression. Engraftment of these edited CD4⁺ T-cells persisted, and as defined by the parameters of the study, the application of gene edited CD4⁺ T-cells in human subjects was concluded as safe (Tebas *et al.*, 2014).

Table 1.1. List of human clinical trials harnessing zinc-finger nuclease to target HIV1 infection

Identifier	Title of study	Phase	Status	Reference
NCT00842634	Autologous T-Cells Genetically Modified at the <i>CCR5</i> Gene by Zinc Finger Nucleases SB-728 for HIV (Zinc-Finger)	1	Completed	(Tebas <i>et al.</i> , 2014)
NCT02388594	A Phase 1 Study of T-Cells Genetically Modified at the <i>CCR5</i> Gene by Zinc-Finger Nucleases SB-728mR in HIV-Infected Patients	1	Active	n/a
NCT01044654	Phase 1 Dose Escalation Study of Autologous T-cells Genetically Modified at the <i>CCR5</i> Gene by Zinc Finger Nucleases in HIV-Infected Patients	1	Completed	n/a
NCT02500849	Safety Study of Zinc Finger Nuclease <i>CCR5</i> -modified Hematopoietic Stem/Progenitor Cells in HIV1 Infected Patients	1	Recruiting	n/a
NCT02225665	Repeat Doses of SB-728mR-T After Cyclophosphamide Conditioning in HIV-Infected Subjects on HAART	1/2	Active	n/a

n/a, not applicable

Substantial progresses in proof-of-concept studies have been made, which demonstrate the advantage of utilising the CRISPR/Cas9 as a gene therapy. Mouse disease models that centre on a single culprit gene mutation have been particularly useful in demonstrating that CRISPR/Cas9 can correct and rescue disease phenotypes. Initial *in vivo* studies examined the effect of delivering CRISPR/Cas9 directly into the animal germline to target a defined mutation; as was exemplified by the mutated *Crygc* gene in mice (Wu *et al.*, 2013). The 1bp deletion in exon 3 of *Crygc* produces a truncated protein product that confers a dominant cataract disorder. Repair of mutant *Crygc* by NHEJ or HR with exogenous DNA encoding the wildtype sequence at the zygote level resulted in rescue from the diseased phenotype in a modest proportion of the P1 generation (approximately 30%) (Wu *et al.*, 2013). Additionally, the repaired *Crygc* gene was transmitted to the next generation (Wu *et al.*, 2013). However, due to ethical concerns of germline manipulation in humans and human embryos, the approach explored by Wu and colleagues (2013) presents limited translatability as a curative human gene therapy. This is even in despite of the fact that between 10-25% cases of congenital cataracts are a result of genetic mutations, which includes the human homologue of the *Crygc* gene – *CRYGC* (Zhong *et al.*, 2017).

Instead, somatic gene editing using CRISPR/Cas9 holds more potential translation as a novel medicine. The application of somatic gene editing presents with fewer ethical considerations. Recent research has successfully shown clinical benefit and gene correction for a number of monogenetic diseases based on animal models. Haemophilia B is an X-linked genetic bleeding disorder attributed by a deficiency in the blood coagulator factor IX, and caused by mutations in the human factor IX (*FIX*) gene (Rallapalli *et al.*, 2013). Reconstituting a small proportion of factor IX levels in the plasma resulted in significant restoration to the clotting activity (Wang *et al.*, 2000a; Nathwani *et al.*, 2014). Perhaps one of the initial-most proof-of-concept, *in vivo* studies to explore the curative potential of CRISPR/Cas9 for haemophilia B was performed by Guan *et al.* (2016). Here, CRISPR/Cas9 and donor template were delivered as naked DNA via the tail veins of haemophilia B mice model, which successfully corrected the factor IX mutation *in situ*. Coagulation activity in the haemophilia B mice was restored and survival rates increased after tail-clip challenges to nearly 90% from 38% of the control group (Guan *et al.*, 2016). The corrected mouse *FIX* gene occurred in approximately 0.56% of endogenous alleles by HR (Guan *et al.*, 2016). This was sufficient to ameliorate the clotting deficiency phenotype. Similar results were reported by additional studies using CRISPR/Cas9 *in situ* to correct haemophilia B mouse models (Huai *et al.*, 2017; Ohmori *et al.*, 2017). However, the use of

mouse *FIX* knockout haemophilia B models function well to recapitulate the severe haemophilia B phenotype in animals, a number of human cases of haemophilia B are a result of dysfunctional factor IX protein being expressed due to single, missense point mutation. This results in the deficiency in functional factor IX expression overall, and not necessarily a complete knockout phenotype *per se* (Rallapalli *et al.*, 2013; Guan *et al.*, 2016; Yuen *et al.*, 2017). Therefore, the impact dysfunctional factor IX has and the interplay this has with corrected factor IX in haemostasis cannot be taken into consideration when applying CRISPR/Cas9 to mouse *FIX* knockout models. Perhaps *in situ* gene therapy of more directly comparable mouse models of human haemophilia B, of which analogous *FIX* mutations are represented, would provide more confident translation potential. More recently, the autologous, *ex vivo* expansion of haemophilia B patient-derived iPSCs was subjected to gene correction by CRISPR/Cas9 genome editing to restore native sequence or factor IX cDNA knock-in into exon 1 of the endogenous *FIX* gene (Ramaswamy *et al.*, 2018). *In vitro* differentiation of corrected iPSC lines into hepatocyte-like cells was followed by transplantation into immune-deficient *FIX* knockout mice. These transplantations correlated with detectable factor IX expression, increased clotting efficiency (rescued to approximately 25% clotting activity of wildtype), and grafts were well sustained (Ramaswamy *et al.*, 2018).

Duchenne muscular dystrophy is characterised by progressive muscle weakening and degeneration, due mutations in the dystrophin gene (*DMD*) and absence of the dystrophin protein (Mital *et al.*, 1998). Restoration of a functional dystrophin protein product was successfully achieved using CRISPR/Cas9 (AAV-CRISPR) to excise the mutated exon from post-natal *mdx* mouse model of Duchenne muscular dystrophy *in vivo*. This resulted in the expression of a truncated form of dystrophin that retained partial wildtype functionality (Long *et al.*, 2016; Nelson *et al.*, 2016; Tabebordbar *et al.*, 2016). The modified dystrophin was successfully found expressed in cardiac and skeletal muscles, and provided partial recovery of muscle function (Long *et al.*, 2016; Nelson *et al.*, 2016; Tabebordbar *et al.*, 2016). Alternatively, Zhu *et al.* (2017) achieved *ex vivo* culture and expansion of muscle stem cells derived from *mdx* mice for HR-mediated genome editing to repair of the mutated *Dmd* gene by Cas9 endonuclease via an Adenovirus vector. Following the transplantation of the gene-corrected muscle stem cells into *mdx* mice, an appreciable level of dystrophin was detectable in skeletal muscle. However, the study fails to inform us on whether or not a clinical benefit was measurable in the *mdx* mice using gene-corrected muscle stem cells. This methodology was in some contrast to the previously proposed *in vivo* approaches that relied on Cas9 endonuclease

function to excise the mutated *Dmd* exon and NHEJ to repair the break and facilitate encoding an in-frame dystrophin protein by exon skipping. The benefit of genome editing muscle stem cells is the capacity for stem cells and replicating cells to mediate HR, which was a presumably lost feature of terminally differentiated cells such as myocytes (Saleh-Gohari and Helleday, 2004). This has recently been contested – non-dividing cardiomyocytes were found to utilise HR, and were capable of being genome edited using CRISPR/Cas9 (Ishizu *et al.*, 2017).

As a result of the promising capacity of CRISPR/Cas9 genome editing *in* animal models, a boom in several, early phase human clinical trials examining the safety profile of *ex vivo* CRISPR/Cas9 genome edited cell lines are to be conducted (Table 1.2). In this regard, the function of CRISPR/Cas9 genome editing is being exploited to enhance the cellular therapy product, rather than directly apply it as a gene therapy, gene correcting tool. As is represented by Table 1.2, a number of human clinical trials will involve the application of CRISPR/Cas9 to disrupt endogenous T-cell receptor (TCR) and/or programmed cell death protein (PD)1 expression from autologous T-cells. The goal is to generate T cells expressing tumour specific TCRs to differentially recognise and target cancer cells *in vivo* (chiefly by clinical trial NCT03399448). PD-1 disruption was warranted due to studies implicating PD-1 expression and its constitutive stimulation of tumour-specific T-cells with impaired T-cell function (Blank *et al.*, 2006; Wei *et al.*, 2013; Wu *et al.*, 2014). This was due to, in part, the PD-1 ligation to the PD-L1 ligand that was expressed by tumour cells in order to coax peripheral immune tolerance and favour its survival (Blank *et al.*, 2006; Wei *et al.*, 2013; Wu *et al.*, 2014). Therefore, the disruption of PD-1 expression from T-cells aims to augment the capacity of tumour-specific T-cells to mediate killing of PD-L1⁺ tumours by eliminating this check-point inhibitory stimulation (Su *et al.*, 2016a; Rupp *et al.*, 2017). Additionally, the existing endogenous TCR effectively competes with the tumour specific TCR for surface expression, and is capable of forming mixed dimers. This may compromise tumour-specificity (van Loenen *et al.*, 2010; Bunse *et al.*, 2014). Elimination of endogenous TCR is thought to generate more efficacious T cell therapy product.

Assessment of the safety profile of using CRISPR/Cas9 genome editing to produce gene and cellular therapy products for human clinical trials is a promising step in the direction of generating a novel class of medicine. The concept of harnessing the genome editing tool to improve or enhance a gene and/or cellular therapy product is a rather novel field of research. It remains, however, to objectively interpret the data from the aforementioned human clinical trials when completed before further progress can be recommended.

Table 1.2. List of human clinical trials utilising CRISPR/Cas9 genome editing

Identifier	Title of study	targeted condition	Phase	Status
NCT03057912	A Safety and Efficacy Study of TALEN and CRISPR/Cas9 in the Treatment of HPV-related Cervical Intraepithelial Neoplasia	HPV-related malignant neoplasm	1	Not yet recruiting
NCT03164135	Safety of Transplantation of CRISPR CCR5 Modified CD34+ Cells in HIV-infected Subjects With Haematological Malignancies	HIV1 infection	Not determined	Recruiting
NCT03399448	NY-ESO-1-redirected CRISPR (TCRendo and PD1) Edited T Cells (NYCE T Cells)	Multiple cancers	1	Recruiting
NCT03081715	PD-1 Knockout Engineered T Cells for Advanced Oesophageal Cancer	Oesophageal cancer	2	Recruiting
NCT02863913	PD-1 Knockout Engineered T Cells for Muscle-invasive Bladder Cancer	Invasive bladder cancer stage IV	1	Not yet recruiting
NCT02867345	PD-1 Knockout Engineered T Cells for Castration Resistant Prostate Cancer	Hormone refractory prostate cancer	1	Not yet recruiting
NCT02867332	PD-1 Knockout Engineered T Cells for Metastatic Renal Cell Carcinoma.	Metastatic renal cell carcinoma	1	Not yet recruiting
NCT02793856	PD-1 Knockout Engineered T Cells for Metastatic Non-small Cell Lung Cancer	Metastatic non-small cell lung cancer	1	Recruiting
NCT03044743	PD-1 Knockout EBV-CTLs for Advanced Stage Epstein-Barr Virus (EBV) Associated Malignancies	Multiple cancers	1/2	Recruiting

1.7.0 Aims and objectives

The study and application of gene therapy products are intensely growing fields of research. Successful gene therapy has been especially exemplified by the successful introduction of AAV-based gene therapy products – Glybera and Luxturna – as the world’s first gene therapy products to be clinically available for the treatment of monogenetic diseases. Therefore, there is considerable promise in gene therapy, especially when mediated by rAAV vector technology, as a means to treat and cure select human diseases. However, the clinical efficacy of using rAAV vectors is restricted by limitations pertaining to our limited understanding of the roles host cellular factors have towards AAV assembly and processing. Such limitations also include a pre-existing immunity towards current AAV serotypes (Kotterman *et al.*, 2015), and the insufficient production of rAAV vector titres to meet the demands of clinical studies.

Advancements in rAAV vector production and technology includes the development of partially stable cell lines for *rep* and/or *cap* expression (Gao *et al.*, 2002b; Mietzsch *et al.*, 2014), or exhibit improved transduction profiles or have enhanced target specificities (Aslanidi *et al.*, 2013; Buning *et al.*, 2015; Ling *et al.*, 2016), and the adaptation of the baculovirus expression system to generate rAAV vectors from insect cells (Urabe *et al.*, 2002; Smith *et al.*, 2009). However, given our limited understanding of the roles host cellular factors have towards AAV assembly and processing, there is a need to exploit this avenue of research for the enhancement of rAAV vector production and rAAV vector-based gene therapy products.

Previously, it has been identified that the endogenous human cell factor, YB1, was incorporated into rAAV vectors, and that gene manipulation of YB1 protein expression by shRNA-mediated knockdown in 293T cells correlated with enhanced rAAV2 and rAAV8 vector titres (Satkunanathan *et al.*, 2014). Therefore, it was postulated that YB1 may have an important role in rAAV vector production and processing. However, a transient rAAV titre enhancement was observed by shRNA-mediated targeting of YB1 protein expression. With these in mind, the aims of this thesis were to i) utilise the CRISPR/Cas9 genome editing system to target the YB1-encoding gene (*YBX1*) in 293T cells to generate a novel YB1 knockout cell line for rAAV vector production; ii) extensively characterise the novel YB1 knockout cell lines at the molecular and phenotypical levels as conducive for rAAV vector production, and establish a well-controlled rAAV production model for effective comparisons; iii) examine the role of wildtype YB1, YB1 truncation mutants and YB1 knockout in rAAV biology and vector production; and lastly, iv) identify the YB1 orthologue and orthologous gene in *S. frugiperda*,

for genome manipulation using CRISPR/Cas9 and subsequent analysis on rAAV vector production. The overall purpose of this project was to examine the utility of CRISPR/Cas9 genome editing in generating a novel rAAV vector production system from both 293T and Sf9 cell lines, and in doing so, examine the impact YB1 and YB1 knockout (or its orthologue in the Sf9 cell) has on rAAV vector production.

It is important to acknowledge and appreciate that the use of CRISPR/Cas9 genome editing has been largely been researched for its direct application in genome editing as a potentially novel method for gene therapy i.e. gene editing to correct genetic defects associated with a disease phenotype. Little research has been performed using CRISPR/Cas9 genome editing to enhance a gene therapy product such as viral vector production platforms (Satkunanathan *et al.*, 2014; Satkunanathan *et al.*, 2017). Core research, instead, focuses on modifying AAV capsids, transgenes and/or expression cassettes, with little attention on the AAV vector producer factory i.e. the cell line itself, to address the restricted vector titres achieved by current platforms. Instead, more focus is centred on delivering CRISPR/Cas9 for somatic genome editing to correct defective genes and clinical phenotype (Bengtsson *et al.*, 2017; Pingjuan *et al.*, 2018). Perhaps the closest examples of modifications of the actual cellular or gene therapy products using CRISPR/Cas9 in the current literature is the generation of engineered T cells with endogenous PD-1 (Su *et al.*, 2016a; Rupp *et al.*, 2017) or TCR (Osborn *et al.*, 2016; Ren *et al.*, 2017) disruption by CRISPR/Cas9 genome editing, *ex vivo*. Here, the cellular therapy product is being modified for improved performance in tumour cell killing rather than using CRISPR/Cas9 for somatic genome editing *in vivo*. Such edited cells are to be utilised in a number of early phase human clinical trials (Table 1.2). Given this encouraging spur in the use of CRISPR/Cas9 genome editing to develop enhanced cellular therapy products, the utilisation of the genome editing technology to potentially produce enhanced rAAV vector production systems is certainly warranted. However, little exploration in endogenous gene manipulation has been performed on producer cells for enhanced rAAV vector titres or production; of which this thesis will be first to represent.

Chapter 2: Materials and Methods

2.1.0 Materials

2.1.1 General laboratory reagents

Common reagents are listed below in Table 2.1, with more protocol-specific reagents described in their relevant sections.

Table 2.1 General reagents used

Reagent	Supplier
β -Mercaptoethanol	Sigma-Aldrich
Acetic acid	VWR
Agarose	Sigma-Aldrich
Ammonium persulphate (APS)	Sigma-Aldrich
Bovine Serum Albumin (BSA)	Sigma-Aldrich
Bromophenol blue	Sigma-Aldrich
Calcium Chloride (CaCl_2)	Sigma-Aldrich
Coomassie [®] Brilliant Blue R-250	Sigma-Aldrich
Dimethyl sulphoxide (DMSO)	Sigma-Aldrich
Dipotassium phosphate (K_2HPO_4)	Sigma-Aldrich
Disodium phosphate (Na_2HPO_4)	Sigma-Aldrich
Dithiothreitol (DTT)	Sigma-Aldrich
Ethanol	France Alcools
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Formaldehyde solution (37-41%)	Fisher Scientific
Glycerol	Sigma-Aldrich
Glycine	Fisher Scientific
HEPES	Sigma-Aldrich
Hydrochloric acid (HCl)	VWR
Isopropanol	Sigma-Aldrich
Magnesium Chloride (MgCl_2)	Sigma-Aldrich
Methanol	VWR

N,N,N',N'-Tetramethylethylenediamine (TEMED)	Sigma-Aldrich
Sodium chloride (NaCl)	Sigma-Aldrich
Sodium deoxycholate	Sigma-Aldrich
Sodium dodecyl-sulphate (SDS)	Sigma-Aldrich
Sodium hydroxide (NaOH)	VWR
Triton™ X-100	Sigma-Aldrich
Trizma base (Tris-base)	Sigma-Aldrich
Tween™ 20	Sigma-Aldrich

2.1.2 Oligonucleotides (oligos)

All polymerase chain reactions (PCR), cloning and sequencing oligos are listed in Table 2.2. Oligos that correspond to gRNA DNA designs for CRISPR/Cas9 genome editing of 293T cells are listed in Table 2.3. Oligos that correspond to gRNA DNA designs for CRISPR/Cas9 genome editing of Sf9 cells are listed in Tables 2.4 and 2.5. Oligos used to generate or function as desthiobiotin- or biotin- labelled DNA capture probes are listed in Table 2.6. Most oligos were synthesised by Sigma-Aldrich, whereas 5'-desthiobiotin or 5'-biotin labelled oligos were synthesised by Integrated DNA Technologies (IDT). All oligos were prepared as 100µM stock solutions in nuclease-free water (H₂O) unless stated otherwise, and further diluted to 10µM working stocks in nuclease-free H₂O. All oligos were stored at -20°C. Where relevant, the applications in which oligos are used and annealing temperatures (T_a) for PCR are also described.

Table 2.2 List of cloning and sequencing oligos

oligo name	5'>3' sequence	Application	T _a (°C)
U6.F(orfward)	GGA CTA TCA TAT GCT TAC CG	Sequencing	n/a
Surv-gRNA1.F(orfward)	GTC GCT CGT AGG GCT TAT CC	Sequencing, PCR	58*
Surv-gRNA1.R(everse)	CTA ACG GTT CCG CTG CTG	(Surveyor [®] assay)	
Surv-gRNA2.F(orfward)	GTC ATT TCT GTG CAC CCC TG	Sequencing, PCR	58
Surv-gRNA2.R(everse)	CTC CAT CAC TTC TCC CTG CA	(Surveyor [®] assay)	
Surv-gRNA3.F(orfward)	ACT CTT GGA TTT ATC TGG T	Sequencing, PCR	50
Surv-gRNA3.R(everse)	TTA TCT TAC CTG TTC AAG GA	(Surveyor [®] assay and HRM template)	
Surv-gRNA4.F(orfward)	TTA CAC AAA TTG CCT ACA GA	PCR	50
Surv-gRNA4.R(everse)	ATA AAT GGA TGG CCA TCC AA		
Surv-sgRNAsf1.F(orfward)	ACT CTA ACC TGA CTT CAC AG	Sequencing, PCR	56
Surv-sgRNAsf1.R(everse)	ACA CAC GAT CAA TAA CGC CG	(Surveyor [®] assay)	
HRM-KOC.F(orfward)	TGA GGC AGA ATA TGT ATC GG	HRM	58
HRM-KOC.R(everse)	CTA ACA CGG GTT TGA GGA AA		
HRM-5'KOC.F(orfward)	CCC AGT AGG CTT AAT TTC CA	HRM	58
HRM-5'KOC.R(everse)	CCG ATA CAT ATT CTG CCT CA		
HRM-3'KOC.F(orfward)	TTT CCT CAA ACC CGT GTT AG	HRM	58
HRM-3'KOC.R(everse)	TCA CCT CTT AGG CAT CTG AT		

SfYB-inf.F(oward)	<u>AAT CAA AGG AGA TAT</u> ACC ATG GCA CA	PCR (In-Fusion® cloning)	58
SfYB-inf.R(everse)	ATG GTG ATG ATG <u>GTG GTG</u> TTA AGC CTG		
pIEx1-ie1prom.F(oward)	CAA GAT CGT GAA CAA CCA AG	Colony PCR	55
pIEx1-ie1term.R(everse)	ACT TAG TGC TCG AGA TCC TC	Colony PCR	55
eGFP.colPCR.R(everse)	TTA CTT GTA CAG CTC GTC CA	Colony PCR	55
T7.F(oward)	TAA TAC GAC TCA CTA TAG G	Sequencing	n/a
YBF1	GCT AGG ATC CGC CAC CAT GAG CAG CGA GGC CGA G	High-fidelity PCR	58°C for YBF1 and YB1R2, YBF1 and YBR2, YBF1 and YBR3, and YBF1 and YBR4 oligo pairs, or 54°C for YBF2 and YBR3, YBF3 and YBR1, and YBF2 and YBR1 oligo pairs
YBF2	GCT AGG ATC CGC CAT GAA GAA GGT CAT CGC AAC G		
YBF3	GCT AGG ATC CGC CAC CAT GGT TCC AGT TCA AGG CAG T		
YBR1	GC AGC GGC CGC TTA CTC AGC CCC GCC CTG		
YBR2	GCT AGC GGC CGC TTA TCT AGG CTG TCT TTG GCG		
YBR3	GCT AGC GGC CGC TTA CAC TTC TCC CTG CAC AGG		
YBR4	GCT AGC GGC CGC CGC TTA ACC AGG ACC TGT AAC ATT TG		
SFFV.F(oward)	CCC AAG GAC CTG AAA TGA CC TGC	Sequencing	n/a
WPRE.R(everse)	GCA GCG TAT CCA CAT AGC G	Sequencing	n/a
CMV.F(oward)	TTC CTA CTT GGC AGT ACA TCT ACG	qPCR	60
CMV.R(everse)	GTC AAT GGG GTG GAG ACT TGG		
CMV.hyd_probe	[6FAM]TGA GTC AAA CCG CTA TCC ACG CCC A[TAM]		

*, 98°C denaturation during PCR thermocycling was required. Underlined nucleotides refer homology arms for cloning by In-Fusion®. HRM, high-resolution melting; PCR, polymerase chain reaction; qPCR, quantitative PCR; T_a, annealing temperature.

Table 2.3 *YBX1*-specific gRNA designs to target *YBX1* gene

name	<i>YBX1</i> -specific gRNA sequence (5' > 3')	<i>YBX1</i> -specific gRNA sequence (5' > 3') + 3' overhang	<i>YBX1</i> intron/exon target
<i>YBX1</i> sgRNA1.F(forward)	ATC GGC GGC GCC TGC CGG CG	ATC GGC GGC GCC TGC CGG CG G TTT T	exon 1
<i>YBX1</i> sgRNA1.R(verse)	CGC CGG CAG GCG CCG CCG AT	CGC CGG CAG GCG CCG CCG AT C GGT G	
<i>YBX1</i> sgRNA2.F(forward)	GTA ATG GCT TTT GTA GGG TG	GTA ATG GCT TTT GTA GGG TG G TTT T	exon 5
<i>YBX1</i> sgRNA2.R(verse)	CAC CCT ACA AAA GCC ATT AC	CAC CCT ACA AAA GCC ATT AC C GGT G	
<i>YBX1</i> sgRNA3.F(forward)	GGA CCA TAC CTG CGG AAT CG	GGA CCA TAC CTG CGG AAT CG G TTT T	intron 6
<i>YBX1</i> sgRNA3.R(verse)	CGA TTC CGC AGG TAT GGT CC	CGA TTC CGC AGG TAT GGT CC C GGT G	
<i>YBX1</i> sgRNA4.F(forward)	CAA AGA CAG CCT AGA GAG GA	CAA AGA CAG CCT AGA GAG GAG G TTT T	exon 7
<i>YBX1</i> sgRNA4.R(verse)	TCC TCT CTA GGC TGT CTT TG	TCC TCT CTA GGC TGT CTT TG C GGT G	

The gRNA oligo specific to PAM-containing target sequence required 5'-GTTTT-3' overhang (**red** sequences). The complementary oligo required 5'-CGGTG-3' overhang (**green** sequences), to facilitate cloning into linearised GeneArt™ CRISPR nuclease vector.

Table 2.4 *SFYB*-specific gRNA designs to target *SFYB* gene

name	<i>SFYB</i> -specific gRNA sequence (5' > 3')	<i>YBX1</i> -specific gRNA sequence (5' > 3') + 3'-overhang	potential <i>SFYB</i> intron/exon target
<i>SFYB</i>sgRNA1.F(forward)	CAC CAT GGC TGA TAC CGA AA	CAC CAT GGC TGA TAC CGA AA G TTT T	exon 1
<i>SFYB</i>sgRNA1.R(verse)	TTT CGG TAT CAG CCA TGG TG	TTT CGG TAT CAG CCA TGG TG C GGT G	
<i>SFYB</i>sgRNA2.F(forward)	GTG GAT ATG GTT TCA TCA AC	GTG GAT ATG GTT TCA TCA AC G TTT T	exon 2
<i>SFYB</i>sgRNA2.R(verse)	GTT GAT GAA ACC ATA TCC AC	GTT GAT GAA ACC ATA TCC AC C GGT G	
<i>SFYB</i>sgRNA3.F(forward)	GTG GAG TTT GCC GTG GTT GC	GTG GAG TTT GCC GTG GTT GC G TTT T	exon 3
<i>SFYB</i>sgRNA3.R(verse)	GCAACCACGGCAAACCTCCAC	GCA ACC ACG GCA AAC TCC AC C GGT G	

The gRNA oligos specific to PAM-containing target sequence required 5'-GTTTT-3' overhang (**red** sequences) on the 3'-end. The complementary oligo required 5'-CGGTG-3' overhang (**blue** sequences) on the 3'-end, to facilitate cloning into linearised GeneArt™ CRISPR nuclease vector.

Table 2.5 *SFYB*-specific gRNA designs to target *SFYB* gene

name	<i>SFYB</i> -specific gRNA sequence (5' > 3')	<i>SFYB</i> -specific gRNA sequence (5' > 3') + T7 or crRNA sequences	potential <i>SFYB</i> intron/exon target
IVT <i>SFYB</i> sgRNA1.F(oward)	CAC CAT GGC TGA TAC CGA AA	TAA TAC GAC TCA CTA TAG CAC CAT GGC TGA TAC CGA	exon 1
IVT <i>SFYB</i> sgRNA1.R(everse)	TTT CGG TAT CAG CCA TGG TG	TTC TAG CTC TAA AAC TTT CGG TAT CAG CCA TGG TG	
IVT <i>SFYB</i> sgRNA2.F(oward)	GTG GAT ATG GTT TCA TCA AC	TAA TAC GAC TCA CTA TAG GTG GAT ATG GTT TCA TCA	exon 2
IVT <i>SFYB</i> sgRNA2.R(everse)	GTT GAT GAA ACC ATA TCC AC	TTC TAG CTC TAA AAC GTT GAT GAA ACC ATA TCC AC	
IVT <i>SFYB</i> sgRNA3.F(oward)	GTG GAG TTT GCC GTG GTT GC	TAA TAC GAC TCA CTA TAG GTG GAG TTT GCC GTG GTT	exon 3
IVT <i>SFYB</i> sgRNA3.R(everse)	GCA ACC ACG GCA AAC TCC AC	TTC TAG CTC TAA AAC GCA ACC ACG GCA AAC TCC AC	

To generate the gRNA DNA templates for IVT of sgRNAs, gRNA designs specific for the PAM containing target sequence strand required oligos with **T7** sequences on 5'-end. The complementary oligo required **crRNA** sequence on 5'-end. Underlined G refers to initial G nt where transcription is initiated from T7 promoter. IVT, *in vitro* transcription.

Table 2.6 List of oligos for DNA-affinity YB1 pulldown

oligo name	modification	5'>3' sequence	Application	T _a (°C)
ITR_YB1.F(orward)	n/a	TGG TAG AGA CGG GGT TTC AC	PCR	58
ITR_YB1.R(verse)	5'-desthiobiotin	GGG AAG AAA GCG AAA GGA GC		
rep_YB1.F(orward)	5'-desthiobiotin	CTT CCG GCT CGT ATG TTG TG	PCR	58
rep_YB1.R(verse)	n/a	CGT GCA TGT GGA AGT AGC TC		
E2A_YB1.F(orward)	n/a	TCT TGT GAC GAG TCT TCT TCG	PCR	58
E2A_YB1.R(verse)	5'-desthiobiotin	TAG GTA GTC GCC ATG CCT TT		
L-ITR_YB1.F(orward)	5'-desthiobiotin	CCC ACC AGC CTT GTC CTA AT	PCR	58
L-ITR_YB1.R(verse)	n/a	CAG AAG GAC AGG GAA GGG AG		
R-ITR_YB1.F(orward)	n/a	GCC TGA TGC GGT ATT TTC TCC	PCR	58
R-ITR_YB1.R(verse)	5'-desthiobiotin	GGG TTG AGT GTT GTT CCA GT		
GCcomp.F(orward)	3'-biotin	GGC GGG GGC GGG GGC GGG GGC GGG TTT TT	Capture probe	n/a
GCcomp.F(orward)	n/a	GGC GGG GGC GGG GGC GGG GGC GGG	Capture probe	n/a
GCcomp.R(verse)	n/a	CCC GCC CCC GCC CCC GCC CCC GCC	Capture probe	n/a
Luciferase.F(orward)	3'-biotin	GCA TAG AAC TGC CTG CGT CAG ATT CTC TTT TT	Capture probe	n/a
Luciferase.R(verse)	n/a	GAG AAT CTG ACG CAG GCA GTT CTA TGC	Capture probe	n/a
ITR(-)sense.F(orward)	n/a	GCA GGA ACC CCT AGT GAT GGA GTT GGC	Capture probe	n/a
ITR(-)sense..R(verse)	3'-biotin	GCC AAC TCC ATC ACT <u>AGG GGT TCC</u> TGC TTT TT	Capture probe	n/a
E2A(-)sense.F(orward)	n/a	CGA GAA GGA GGA CAG CCT AAC CGC CCC	Capture probe	n/a

E2A(-)sense.R(everse)	3'-biotin	GGG GCG GTT AGG CT <u>GTC CTC CTT</u> CTC GTT TTT	Capture probe	n/a
rep(+)sense.F(oward)	3'-biotin	GCG CAG CCG CCA TGC <u>CGG GGT</u> TTT ACG TTT TT	Capture probe	n/a
rep(+)sense.R(everse)	n/a	CGT AAA ACC CCG GCA TGG CGG CTG CGC	Capture probe	n/a

Underlined nucleotides refer to YB1 binding motifs (GGGGTT or CCTCCT). n/a, not applicable; PCR, polymerase chain reaction; T_a, annealing temperature.

2.1.3 Chemically competent bacteria – strains

α -Select Silver Efficiency Chemically Competent Cells, *E. coli* (Bioline); genotype: F⁻ *deoR endA1 recA1 relA1 gyrA96 hsdR17(r_k⁻, m_k⁺) supE44 thi-1 phoA Δ(lacZYA argF)U169 φ80lacZΔM15λ⁻*

MAX Efficiency™ DH10Bac™ Competent Cells, *E. coli* (Gibco); genotype: F⁻ *mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻rpsL nupG/pMON14272/pMON7124*

One Shot™ TOP10™, Chemically Competent, *E. coli* (Invitrogen); genotype: F⁻ *mcrA Δ(mmr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU rpsL (Str^R) endA1 nupGλ⁻*

2.1.4 Antibodies

Primary antibodies are listed in Table 2.7. Secondary antibodies are listed in Table 2.8. The applications in which antibodies were used and the corresponding dilution factors are also described.

Table 2.7 List of primary antibodies

primary antibody	listed name	antigen	immunogen	raised in	clonality	manufacturer	application (dilution)
A20	A20	AAV2 intact capsid	unspecified	Ms	mAb	Progen	ICC (1:20)
αCRISPR-Cas9	7A9-3A3	Cas9	unspecified	Ms	mAb	Abcam	WB (5μg/mL)
αGAPDH	CB1001	GAPDH	unspecified	Ms	mAb	Calbiochem	WB (1:1250)
αYB1	ab114999	YB1	synthetic peptide corresponding to 224-274aa of human YB1 (NP_004550)	Rb	pAb	Abcam	WB (1:2000) ICC (1:500)
αYB1	ab76149	YB1	unspecified	Rb	pAb	Abcam	WB (1:10,000) ICC (1:500)
αYB1	PA5-19453	YB1	synthetic peptide corresponding to 1-100aa of human YB1 (NP_004550)	Rb	pAb	Thermo Fisher Scientific	WB (1.4μg/mL) ICC (1:500)
αSOX13	ab96776	SOX13	synthetic peptide corresponding to 75-490aa of human SOX13 (NP_005677)	Rb	pAb	Abcam	WB (1:1000)

αPANK4	ab137243	PANK4	synthetic peptide corresponding to 50-100aa of human pank4 (NP_060686.1)	Rb	pAb	Abcam	WB (1:2000)
αSDF4	ab113413	SDF4	synthetic peptide corresponding to 161-175aa of human SDF4 (NP_057631.1; NP_057260.2)	Gt	pAb	Abcam	WB (0.5μg/mL)
αSfYB	n/a	rSfYB(His) ₁₀	whole protein, 10xHis-tagged recombinant SfYB	Rb	pAb	Covalab	Dot blot (as indicated) WB (1:32,000) ICC (1:32,000)
αRep	10R-A140A	AAV2 Rep	171aa N-terminally truncated AAV2 Rep78	Ms	mAb	Fitzgerald	WB (1:50) ICC (1:50)
αCap	B1	AAV2 Vp1-3	unspecified	Ms	mAb	Progen	ICC (1:100)
αCap	VP51	AAV2 Vp1-3	unspecified	Rb	pAb	Progen	WB (1:200)
α6His-HRP	ab1187	Hexa-His tag	synthetic peptide corresponding to six histadine	Rb	pAb	Abcam	Dot blot (as indicated) WB (1:5000) FC (1:500)

FC, flow cytometry; Gt, Goat; ICC, immunocytochemistry; mAb, monoclonal; Ms, mouse; n/a, not applicable; pAb, polyclonal; Rb, rabbit; WB, western blotting.

Table 2.8 List of secondary antibodies

secondary antibody	conjugate	raised in	manufacturer	Application (dilution)
α Mouse IgG (H+L)	HRP	Sh	Sigma-Aldrich	WB (1:2000)
α Rabbit IgG (H+L)	HRP	Gt	Sigma-Aldrich	WB (1:2000)
α Goat IgG (H+L)	HRP	Rb	Dako	WB (1:2000)
α Rabbit IgG (H+L)	Alexa Fluor 647	Gt	Thermo Fisher Scientific	FC (1:500)
α Mouse IgG (H+L)	DyLight 633	Gt	Thermo Fisher Scientific	ICC (1:500)
α Rabbit IgG (H+L)	Alexa Fluor 547	Gt	Thermo Fisher Scientific	ICC (1:500)
α Rabbit IgG (H+L)	Alexa Fluor 488	Gt	Thermo Fisher Scientific	ICC (1:500) FC (1:500)

FC, flow cytometry; Gt, goat; HRP, horse radish peroxidase; ICC, immunocytochemistry; Rb, rabbit; Sh, sheep; WB, western blotting.

2.1.5 Eukaryotic cell lines

Mammalian and insect cell lines used are detailed in Table 2.9.

Table 2.9 List and details of cell lines.

cell line	organism (origin)	organ origin	characteristics	Supplier
293T	<i>H. sapiens</i>	kidney	Derived from 293 cells that were transformed by introducing sheared Adenovirus 5 DNA containing the <i>E1</i> gene (Graham <i>et al.</i> , 1977), and further transformed by transfection of 293 cells with plasmid encoding temperature-sensitive mutant of SV40 large T antigen and neomycin resistance (DuBridge <i>et al.</i> , 1987).	Stratagene
Ao38	<i>T. ni</i>	ovaries	Derived from single cell cloning by serial dilution in 96-well plates of primary culture of ovarian tissue harvested from <i>T. ni</i> (Hashimoto <i>et al.</i> , 2010; Hashimoto <i>et al.</i> , 2012).	A kind gift from Prof. Ian Jones, University of Reading

Sf9	<i>S. frugiperda</i>	ovaries	Sf9 cell line was cloned from the parent line, IPLB-SF21-AE (Sf21), which was derived from pupal ovarian tissue of <i>S. frugiperda</i> by Vaughn <i>et al.</i> (1977).	Invitrogen
Sf-9ET	<i>S. frugiperda</i>	ovaries	Derived from Sf9 cell line, transfected with plasmid encoding <i>EGFP</i> gene under promoter control of <i>polh</i> , and neomycin resistance gene for selection (Hopkins and Esposito, 2009).	A kind gift from Prof. Ian Jones, University of Reading
Sf9-disrupted	<i>S. frugiperda</i>	ovaries	Derived from Sf9 cells, genome engineered for targeted knockout of SfYB protein expression.	This study
YB1 knockout	<i>H. sapiens</i>	kidney	Derived from 293T cells, genome engineered for targeted knockout of YB1 protein expression.	This study

2.2.0 Methods

2.2.1 Molecular cloning

2.2.1.1 Designing single guide (sg)RNAs for CRISPR/Cas9

CRISPR/Cas9 sgRNA designs against human *YBX1* were designed using the online CHOPCHOP tool (<http://chopchop.cbu.uib.no/>) (Montague *et al.*, 2014), and four separate sgRNAs targeting exon 1, 5, 7, and intron 6 were selected. CRISPR/Cas9 sgRNA designs specific against exons 1-3 of the putatively identified *Spodoptera frugiperda* Y-Box protein homologue gene, after construction of its gene structure (see section 2.2.2.8), were designed manually, instead. Generally, sgRNA designs were 20nt in length and directly upstream of the 5'-NGG-3' PAM sequence on the target DNA strand, to facilitate Cas9 recognition and endonuclease activity 3nt upstream of the PAM (Anders *et al.*, 2014). The CHOPCHOP online tool additionally provides efficiency scores of sgRNA designs that are normalised between 0-1, of which algorithm takes into factor potential off-target sites as a result of 1, 2, or 3 mismatches, and the identification of a guanine residue at sgRNA position 20, just upstream of the PAM (Doench *et al.*, 2014; Xu *et al.*, 2015; Doench *et al.*, 2016). The sgRNAs were ordered as ssDNA oligos corresponding to that target, PAM-containing sequence (forward oligo) and the complementary sequence (reverse oligo), with the appropriate 5'-overhangs for direct cloning into the GeneArt™ linearised, all-in-one CRISPR nuclease vector (forward oligo: 5'-GTTTT-3',

and reverse oligo: 5'-CGGTG-3'), or for assembly PCR and *in vitro* transcription (IVT) (forward oligo: 5'-T7 promoter sequence, and reverse oligo: 5'-crRNA sequence). See Tables 2.3-5 for sgRNA designs.

2.2.1.2 Standard polymerase chain reaction (PCR)

Standard PCRs were used to amplify DNA sequences of interest. Therefore, in order to perform PCRs, forward (corresponding to the sequence in the sense strand) and reverse (corresponding to the sequence in the anti-sense strand) oligos were designed and ordered). In the case of PCR amplification of sequences intended for cloning, forward and reverse oligos would additionally contain the appropriate sequences required for restriction enzyme digestion plus buffer DNA bases on their 5'-ends.

PCRs were generally performed as 50µL reactions, and comprised of the following at their final concentrations: 1X Master Mix (Promega) constituting 25U/mL *Taq* polymerase, 200µM of each dNTP (dATP, dGTP, dCTP, and dTTP), and 1.5mM MgCl₂, forward and reverse oligos at 0.5µM each, and 50ng of genomic DNA or 10ng of plasmid DNA as template, made to 50µL with nuclease-free H₂O. Non-template control PCR reactions were prepared in parallel with template DNA substituted for nuclease-free H₂O. PCRs were performed on a Biorad PTC-200 DNA Engine Thermal Cycler with the programmed thermocycling conditions: 95°C initial denaturation for 2mins, followed by 34 cycles of denaturing at 95°C (unless stated otherwise) for 1min, annealing at a suitable temperature (dependent on the oligo pairs) for 30secs, with an extension step at 72°C for 1min/kb of DNA to be amplified. A final extension step at 72°C for 5mins followed, and reactions were stored at 4°C.

2.2.1.3 Colony PCR

Colony PCR was performed to screen transformed bacteria for correct cloning of inserts using REDTaq® ReadyMix PCR Reaction Mix (Sigma-Aldrich). Colony PCRs were set up as 25µL reactions, comprising of the following at their final concentrations: 1X REDTaq® ReadyMix PCR Reaction Mix, constituting 30U/mL *Taq* polymerase, 200µM of each dNTP (dATP, dGTP, dCTP, and dTTP), 10mM Tris-HCl, 50mM KCl, 3mM MgCl₂, 0.001% gelatin, stabilisers, forward and reverse oligos at 0.4µM each, mixed with a picked transformed bacteria colony (using sterile pipette tip), made to 25µL with nuclease-free H₂O. Non-template control reactions were also run in parallel, and involved the aforementioned setup without adding transformed bacteria colony. PCR reactions were performed on a Biorad PTC-200 DNA Engine Thermal Cycler, with programmed thermocycling conditions: 94°C initial denaturation for 2mins, followed by 34

cycles of denaturation at 94°C for 1min, annealing of oligos at 55°C for 2mins, and extension at 72°C for 3mins. Final extension at 72°C for 5mins followed, after which, reactions were held or stored at 4°C.

2.2.1.4 High-fidelity PCR

High-fidelity PCR was utilised to PCR amplify target sequences with high fidelity and accuracy for cloning. High-fidelity PCR was performed using Q5[®] Hot Start High-Fidelity 2X Master Mix (NEB). Reactions were set up by mixing 10ng of plasmid template with Q5[®] Hot Start High-Fidelity 2X Master Mix made to a final 1X concentration, constituting 200µM of each dNTP (dATP, dGTP, dCTP, and dTTP), 2mM MgCl₂, and final concentration of 0.5µM of each forward and reverse oligo, made to 50µL using nuclease-free H₂O. Thermocycling conditions were programmed on a Biorad PTC-200 DNA Engine Thermal Cycler: initial denaturation at 98°C for 10mins; followed by 34 cycles of denaturation at 98°C for 1min, annealing step for 1min at the optimal annealing temperature (depending on the oligo pairs), extension at 72°C for 1min; and a final extension step at 72°C for 2mins. Reactions were then stored at 4°C.

2.2.1.5 Assembly PCR

Assembly PCR was utilised to produce the gRNA DNA template for *in vitro* transcription (IVT). Assembly PCR functioned to assemble the target-specific forward and reverse assembly oligonucleotides (see Table 2.5 for IVT oligo pairs) with Tracr fragment plus universal T7 primer mix as per GeneArt™ Precision gRNA Synthesis Kit (Invitrogen). IVT-designated forward and reverse oligos included the required 5'-T7 or crRNA sequences, respectively, as per manufacturer's guidelines (Invitrogen). Firstly, 0.3µM oligo mix working solution was prepared by taking 3µL of each 10µM oligo stock for each oligo pair, and diluting in 94µL of nuclease-free H₂O. This was then briefly vortexed and spun down. 1µL of this 0.3µM mix was used to set up PCR reactions constituting Phusion™ High-Fidelity PCR Master Mix (final 1X concentration), 1µL of Tracr Fragment+T7 Primer Mix, made to 25µL using nuclease-free H₂O. Non-template control reaction was included, which included the exclusion of the Tracr Fragment+T7 Primer Mix template, with nuclease-free H₂O substituted instead. PCR thermocycling conditions followed manufacturer's guidelines, and was programmed using a Biorad PTC-200 DNA Engine Thermal Cycler: initial denaturation at 98°C for 10secs; followed by 32 cycles of denaturation at 98°C for 5secs, annealing at 55°C for 15secs; and a final extension at 72°C for 1min. Reactions were held at 4°C and 5µL samples mixed with 6X DNA Loading Dye (Thermo Fisher

Scientific) for subsequent 2% TBE agarose gel electrophoresis to determine PCR assembly reactions.

2.2.1.6 *In vitro* transcription (IVT)

Assembly PCR products were then subjected to IVT as per GeneArt™ Precision gRNA Synthesis Kit (Invitrogen) to generate the sgRNA for CRISPR/Cas9 genome editing applications. 6µL of the gRNA DNA template (generated by assembly PCR) was mixed with 2µL of TranscriptAid™ Enzyme Mix and NTP mix (ATP, CTP, GTP, and UTP) for a final 40mM concentration, 5X TranscriptAid™ Reaction Buffer made to 1X, to a final volume of 20µL. The IVT reaction was incubated at 37°C for 3h. This was followed by the addition of 1U DNase I (Invitrogen) and incubated at 37°C for 15mins to degrade the DNA template. The sgRNA was then purified by adjusting the volumes of IVT reactions to 200µL in nuclease-free H₂O, followed by the addition of 100µL of IVT Binding Buffer (composition – propriety information) and 300µL of 100% ethanol. Mixtures were mixed by pipetting and then transferred to GenJET™ RNA Purification Micro Columns and centrifuged at 13,000rpm for 1min. The flow-through was discarded and the column washed with 700µL of Wash Buffer 1 (composition – propriety information; diluted with 13mL 100% ethanol prior to use) and centrifuged at 13,000rpm for 1min. The flow-through was discarded and the column washed twice with 700µL of Wash Buffer 2 (composition – propriety information; diluted with 30mL 100% ethanol prior to use) and centrifuged at 13,000rpm for 1min. The flow-through was discarded and the column centrifuged at 13,000rpm for 1min to remove the residual Wash Buffer 2. The column was then transferred to a clean 1.5mL Collection Tube, and 20µL of nuclease-free H₂O was added to the centre of the column. This was followed by centrifugation at 13,000rpm for 1min to elute the sgRNA. The sgRNA was quantified by Qubit™ 3.0 Fluorometer (Thermo Fisher Scientific) and Qubit™ RNA BR Assay Kit (Invitrogen) and stored at -20°C. IVT prep quality was determined by mixing 0.5µL of IVT reaction or purified product in 10µL of DEPC-treated H₂O (Sigma-Aldrich), and mixed with 6X DNA Loading Dye. Samples were heated at 70°C for 10mins, and then chilled on ice prior to 2% TBE agarose gel electrophoresis.

2.2.1.7 Restriction enzyme digests

For analytical restriction enzyme digestions, 1µg of plasmid was digested with 5U of enzyme/s as 20µL or 50µL reactions for NEB- or Promega- based reactions, respectively. Digestions were performed in final 1X concentration of suitable buffer that was supplied by the manufacturer, and made to the corresponding volumes in ultra-pure H₂O. For the purpose of generating cloning fragments, 30U of restriction enzyme(s) was used to digest 8µg of plasmid DNA.

Reactions were performed as 100µL reactions in the appropriate buffer supplied by the manufacturer, made to 1X using ultra-pure H₂O. For reactions with two restriction enzymes, a compatible buffer shared by both enzymes (as indicated by the manufacturer) was used. Where appropriate, undigested plasmid control(s) were included and involved omission of the restriction enzyme(s) to reaction mixes, with ultra-pure H₂O substituted instead. Restriction enzyme digestions were performed at the indicated temperature (as indicated by the manufacturer) for the given restriction enzyme(s) used – typically 37°C for 1h (3h for cloning fragments). Restriction enzyme digests were then heat-inactivated at the required temperature and incubation time as indicated by the manufacturer. After which, digestion reactions intended for the isolation of DNA fragments were subjected to dephosphorylation of DNA ends (where indicated), and/or 1% TAE agarose gel electrophoresis for gel extraction.

2.2.1.8 Annealing complementary pairs of oligos

Annealing of complementary pairs of oligos was performed by mixing 50µL of complementary pairs of oligos (100µM stock) together. The mixture was then heated to 95°C for 5mins to denature oligos. Following which, the oligo mix was allowed to cool down slowly to room temperature (RT) for approximately 2h. However, DNA oligo pairs that corresponded to gRNA designs (listed in Tables 2.3 and 2.4) were annealed following the GeneAart® CRISPR Nuclease Vector Kit (Invitrogen) instructions. 200µM oligo stock solutions were prepared using nuclease-free H₂O. 5µL of each 200µM oligo pair were mixed together in a final 1X Oligonucleotide Annealing Buffer (10X: 100mM Tris-HCl [pH 8.0], 10mM EDTA [pH 8.0], 1M NaCl), made up to 20µL with nuclease-free H₂O. The mixture was briefly spun down and 5µL was transferred to a clean Eppendorf tube, followed by incubation at 95°C for 4mins. Denatured oligos were cooled to RT for 10mins, and then briefly spun down. Annealed oligos was then diluted 1:100 in nuclease-free H₂O to 500nM stock solution, which was briefly vortexed and spun down. 5nM working solution was then prepared by diluting 500nM stock solution 1:100 in a final 1X Oligonucleotide Annealing Buffer, made to 100µL with nuclease-free H₂O. The working stock was briefly vortexed and then spun down. 5nM stock was immediately used to ligate into GeneArt™ linearised, all-in-one, CRISPR nuclease vector (Invitrogen).

Alternatively, 3'-biotin labelled DNA capture probes were annealed by mixing 5µL of each 10µM oligo pairs (Table 2.6) to a final volume of 100µL in Binding Buffer A (25mM HEPES [pH 7.5], 100mM KCl, 12.5mM MgCl₂, 1mM DTT, 0.1% IGEPAL-CA630 [v/v], 20% glycerol [v/v], 3%

BSA). The cold competitor DNA probe was prepared separately for each intended ratio of hot:cold DNA capture probe ratio, by scaling up or down the volume of 10 μ M oligo pairs made to a final volume of 100 μ L of Binding Buffer A. Mixtures were then subjected to 55°C for 10mins, followed by cooling to RT for approximately 10mins, and then kept on ice.

2.2.1.9 Dephosphorylating 5'-ends of DNA

To prevent re-linearisation of single restriction enzyme digested plasmids and improve cloning efficiencies, dephosphorylation of 5'-ends of restriction enzyme digested DNA was performed where indicated. Therefore, recombinant Shrimp Alkaline Phosphatase (rSAP; NEB) was added after heat-inactivation of restriction enzyme(s) at 1U/pmol of DNA ends. Dephosphorylation reactions were performed at 37°C for 1h, followed by heat-inactivation at 65°C for 10mins. Restriction enzyme digested and dephosphorylated DNA was then subjected to 1% TAE agarose gel electrophoresis for gel extraction.

2.2.1.10 Blunting DNA ends by Klenow Fragment

DNA ends were blunted using Klenow Fragment (NEB), which removes 3'-overhangs or fill in 5'-overhangs. Approximately 1 μ g of restriction enzyme-digested DNA was mixed in Cutsmart buffer (NEB) to a final 1X concentration, 1U of Klenow Fragment, and supplemented with final concentration of 33 μ M dNTP mix (Promega) to a final volume of 50 μ L using nuclease-free H₂O. This was followed by incubation at RT for 15mins. The Klenow Fragment was then inactivated by the addition of 500mM EDTA (pH 8.0) solution to a final 10mM concentration, and heat-inactivated at 75°C for 20mins. Blunted DNA fragments were stored at -20°C.

2.2.1.11 Phosphorylating DNA ends

Phosphorylation of DNA 5'-ends was performed using T4 PNK (NEB) for subsequent ligation reactions. 300pmol of 5'-ends was calculated for DNA plasmid required for phosphorylation. This was mixed into T4 DNA Ligase Buffer (NEB) to a final 1X (10X: 500mM Tris-HCl, 100mM MgCl₂, 10mM ATP, 100mM DTT, pH 7.5) and 10U of T4 PNK. Phosphorylation reactions were incubated at 37°C for 30mins, and then heat-inactivated at 65°C for 20mins. Phosphorylated DNA fragments were stored at -20°C.

2.2.1.12 Agarose gel electrophoresis

DNA PCR products and fragments or RNA products were subjected to agarose gel electrophoresis to separate DNA or RNA molecules relative to their size. Agarose powder

(Sigma-Aldrich) was dissolved in 1X TAE (50X: 2M Tris-base, 0.95M acetic acid, 50mM EDTA; diluted 1:50 in ultra-pure H₂O), or 1X TBE Buffer (10X: 0.9M Tris, 0.9M boric acid, 25mM EDTA; diluted 1:10 in ultra-pure H₂O), to make final 1% or 2% concentration of agarose, where indicated. 1% agarose gels were generally used for the separation of DNA fragments required for cloning and screening. 2% agarose gels were prepared for small DNA or RNA products. The agarose mixture was dissolved by boiling using a microwave. The mixture was briefly cooled to allow addition of the SYBR™ Safe DNA Gel Stain (Invitrogen) at 1:10,000 dilution and mixed by swirling. The mixture was then poured into a gel tray with a gel comb to generate the gel's sample wells. The gel was then allowed to set at RT, and then transferred to an electrophoresis tank (Bio-rad) and completely immersed in 1X TAE or 1X TBE Buffer for the corresponding gels. Nucleic acid samples were mixed with 6X DNA Loading and loaded into submerged wells with a pre-stained DNA ladder(s) in parallel. Pre-stained DNA ladders included 100bp DNA Ladder (NEB) or GeneRuler 1kb DNA Ladder (Thermo Fisher Scientific), and functioned as reference markers for DNA resolution through an agarose gel for approximate DNA sizing. Gels were run at 80-120V for a suitable length of time (typically up to 1h) to allow for the required separation. DNA fragments required for cloning, however, were run for 3h using a large electrophoresis apparatus. Resolved DNA/RNA molecules designated for screening were visualised and imaged using a UV trans-illuminator (BioDoc-It™ Imaging System; UVP), whereas, DNA fragments required for cloning were excised from gels visualised by a Safe Imager™ 2.0 Blue Light Trans-illuminator (Thermo Fisher Scientific).

2.2.1.13 Gel extraction and purification of DNA fragments

DNA fragments were extracted and purified by first resolving DNA fragments by 1% TAE agarose gel electrophoresis. Excised DNA within agarose gel slices were purified using either MinElute Gel Extraction Kit (Qiagen) or QIAquick Gel Extraction Kit (Qiagen). Mainly, the QIAquick Gel Extraction Kit was used for gel purification of DNA unless otherwise stated. Excised DNA fragment in gel slice was incubated in three volumes of Buffer QG (solubilisation buffer, composition – proprietary information) to one volume of gel, at 50°C until gel slice had completely solubilised, with intermittent vortexing. Following which, one gel volume of RT isopropanol was added and mixed. Mixtures were then transferred to a QIAquick column or MinElute column, for QIAquick Gel Extraction Kit and MinElute Gel Extraction Kit, respectively. Samples were then centrifuged for 1min at 13,000rpm. The flow-through was discarded, and columns were washed by the addition of 750µL of Buffer PE (10mM Tris-HCl pH 7.5, 80% ethanol), and centrifuged for 1min at 13,000rpm. The flow-through was discarded and column

centrifuged again at 13,000rpm for 1min to remove residual ethanol. The column was then placed in a clean Eppendorf tube and 10 μ L or 50 μ L of nuclease-free H₂O (for QIAquick Gel Extraction Kit and MinElute Gel Extraction Kit, respectively) was added to the column membrane, and left to stand at RT for 1min. DNA was eluted by centrifuging columns for 1min at 13,000rpm. Extracted DNA fragments were stored at -20°C.

2.2.1.14 Purification of DNA fragments

PCR purification of DNA products using QIAquick PCR Purification Kit (Qiagen) was performed where gel purification was not recommended. This involved adding five times volume of Buffer PB (composition – proprietary information) followed by the addition of 10 μ L of 3M Sodium acetate (pH 5.0) to the PCR sample. The DNA sample was then loaded into a QIAquick column to bind DNA to the resin by centrifuging columns for 1min at 13,000rpm, with flow-through discarded. The column was then washed with 750 μ L of PE Buffer and centrifuged for 1min at 13,000rpm, and flow-through discarded. Residual PE Buffer was removed by an additional spin at 13,000rpm for 1min, and flow-through discarded. The column was then transferred to a clean Eppendorf tube and 50 μ L of nuclease-free H₂O was applied to the membrane and left to stand for 1min, RT. The DNA was finally eluted by centrifuging columns for 1min at 13,000rpm. Purified DNA was stored at -20°C.

2.2.1.15 DNA ligations

DNA ligations were prepared as 20 μ L reactions, and were performed using T4 DNA Ligase (NEB or Invitrogen, where indicated). Purified vector DNA and restriction enzyme digested DNA fragments were mixed at 3:1 molar ratio of insert to vector. Reaction volumes were made to 20 μ L with nuclease-free H₂O and a final 1X concentration of T4 DNA Ligase Buffer (NEB) and 1U of T4 DNA Ligase (NEB). Reactions were performed at 4°C overnight. Ligations between 2 μ L of annealed gRNA DNA (5nM) and 2 μ L of pre-linearised GeneArt™ CRISPR nuclease vector was performed using 1U of T4 DNA Ligase (Invitrogen), with final concentration of 1X Ligation Buffer (5X: 250mM Tris-HCl [pH7.6], 50mM MgCl₂, 5mM ATP, 5mM DTT, 25% w/v polyethylene glycon-8000), made to 20 μ L with nuclease-free H₂O. Reactions were performed at RT for 10mins.

2.2.1.16 Cloning by In-Fusion®

Cloning by In-Fusion® was performed using the In-Fusion® HD Cloning Kit (Clontech). This involved the directional cloning of a PCR amplicon into the desired vector backbone using the

In-Fusion® enzyme that catalyses the recombination between homology arms that are present at both ends of cloning insert and linearised vector backbone. The desired cloning insert was PCR amplified using In-Fusion® designated oligos (Table 2.2), followed by gel purification. The resultant PCR amplicon contained the homology arms (denoted in Table 2.2 for the relevant primers) complementary to homology arms present in intended, pre-linearised cloning vector. 100ng of purified cloning insert was mixed with 100ng of linearised cloning vector in a final 1X In-Fusion® HD Enzyme Premix (5X: composition – propriety information), made to 10µL using nuclease-free H₂O. The In-Fusion® reaction was incubated at 50°C for 15mins, and then placed on ice to terminate the reaction.

2.2.1.17 Culture and maintenance of bacteria stocks

Cultivation of transformed bacteria requires positive selection screening using antibiotics, given plasmids typically harbour an antibiotic resistance gene(s). Therefore, ampicillin (1000X) stocks were prepared by dissolving ampicillin sodium salt (Sigma-Aldrich) in ultra-pure H₂O for 100mg/mL concentration. Kanamycin (1000X) was prepared by dissolving kanamycin sulphate (Sigma-Aldrich) into ultra-pure H₂O to make 50mg/mL. Tetracycline (1000X) was prepared by dissolving tetracycline hydrochloride (Sigma-Aldrich) in 100% ethanol to make 10mg/mL. Lastly, gentamicin was provided as a 50mg/mL solution (Gibco). Ampicillin, kanamycin and tetracycline stock solutions were all filtered using 0.22µm filters, and stored at -20°C as 1mL aliquots.

Transformed bacteria were cultivated to amplify plasmid DNA, and performed as liquid cultures using Luria Broth (LB) composed of 1% tryptone (Oxoid), 0.5% yeast extract (Oxoid), 0.8% NaCl, pH 7.0. LB agar plates were used for solid phase growth, and involved preparing LB medium as above with the addition of agar powder (Sigma-Aldrich) for a final 1.5% concentration (w/v). Both LB and LB agar mixtures were autoclaved at 122°C. Appropriate antibiotic(s) was added to LB medium for 50µg/mL concentration of ampicillin (1:2000 dilution). However, transformed DH10Bac™ cells were cultivated in LB medium supplemented with kanamycin (final 50µg/mL; 1:1000 dilution), gentamicin (final 7µg/mL; 1:7142 dilution), and tetracycline (final 10µg/mL; 1:1000 dilution). To prepare LB agar plates, solid LB agar was melted and allowed to cool to approximately 55°C to add antibiotic(s) to the appropriate concentration - 100µg/mL ampicillin, or 50µg/mL kanamycin, 7µg/mL gentamicin, and 10µg/mL tetracycline. After which, liquid LB agar with antibiotics was then poured into Sterilin™ 10cm petri-dishes until solidified.

Stocks of transformed bacteria were stored as glycerol stocks, which involved cultivating transformed bacteria as liquid cultures overnight at 37°C, 200rpm. Glycerol stocks were then prepared by mixing bacteria culture with autoclaved 50% glycerol (v/v), to make 20% glycerol stocks. Bacteria glycerol stocks were then immediately stored as 1mL aliquots at -80°C.

2.2.1.18 Transformations

Transformation of chemically competent *E. coli* cells with plasmid DNA was performed to generate bacteria that could cultivate plasmid DNA for amplification. Generally, plasmids or ligation reactions were transformed into α -Select Chemically Competent Cells (Bioline). The α -Select Chemically Competent Cells were removed from -80°C storage and thawed slowly on ice, after which, cells were transferred to pre-chilled tubes as 50 μ L aliquots. 5 μ L of ligation reaction or approximately 100ng of plasmid DNA was transformed into 50 μ L of α -Select Chemically Competent Cells, followed by flicking to mix and 30mins incubation on ice. Cells were then subjected to heat-shock at 42°C for 45secs. Tubes were returned to ice for 2mins, followed by dilution of the transformation reaction with 950 μ L of S.O.C. Medium (Invitrogen). OneShot™ TOP10™ Chemically Competent *E. coli* (Invitrogen) was used for transformation of GeneArt™ CRISPR nuclease ligation reactions. Transformation of TOP10™ *E. coli* was performed similarly as described for α -Select Chemically Competent Cells, except 3 μ L of GeneArt™ CRISPR nuclease ligation reaction was used. In both cases, diluted transformation reactions were incubated for 1h at 37°C, 200rpm, after which, 100 μ L transformation reaction was spread on LB agar plates, supplemented with 100 μ g/mL ampicillin. Plates were incubated overnight at 37°C for the growth of bacterial colonies and positive transformants.

MAX Efficiency™ DH10Bac™ Competent Cells were used to transform pFastBac constructs. DH10Bac™ Competent Cells were removed from -80°C storage and thawed on ice. Cells were then transferred into pre-chilled 15mL falcon tubes, and 1ng of pFastBac plasmid was added and mixed by pipetting up and down. Mixtures were then incubated on ice for 30mins, followed by heat-shock at 42°C for 45secs. Transformation reactions were then returned to ice for 2mins, followed by the addition of 900 μ L S.O.C. Medium. Transformations were incubated at 37°C at 225rpm for 4h. Serial dilutions of transformation reactions were then prepared at 10⁻¹, 10⁻² and 10⁻³ using S.O.C. Medium. LB agar plates containing 50 μ g/mL kanamycin, 7 μ g/mL gentamicin, 10 μ g/mL tetracycline were pre-warmed at 37°C, followed by the addition of 40 μ L of Blue-White Select™ Screening Reagent to LB agar plates and spread evenly. Plates and Blue-

White Select™ Screening Reagent were allowed dry for 15mins at RT prior to the addition of 100µL DH10Bac™ transformed cells. Plates were incubated for 48h at 37°C until white colonies formed. White colonies were re-streaked onto LB agar plates containing 50µg/mL kanamycin, 7µg/mL gentamicin, 10µg/mL tetracycline, with 40µL of Blue-White Select™ Screening Reagent, and incubated for a further 24h at 37°C to isolate single white colonies with more confidence.

2.2.1.19 Plasmid amplification and isolation – mini-preps

5mL LB media (with appropriate antibiotic) was inoculated with approximately 10µL bacteria glycerol stock or a single colony of transformed bacteria. Inoculations were grown overnight at 37°C, 200rpm. The following day, 5mL cultures were transferred to Eppendorf tubes by pelleting cells at 4000rpm for 1min. Plasmid DNA was then extracted using QIAprep Spin Mini-prep Kit (Qiagen). Pelleted bacteria cells were resuspended in 250µL P1 Buffer (50mM Tris-HCl [pH 8.0], 10mM EDTA, 100µg/mL RNase A), followed by the addition of 250µL P2 Lysis Buffer (200mM NaOH, 1% SDS [w/v], including LyseBlue reagent – 1:1000 dilution), which was mixed by inversion six-times and incubated for up to 5mins. Lysis reactions were neutralised by the addition of 350µL N3 Buffer (4.2M Gu-HCl, 0.9M potassium acetate, pH 4.8) and mixed by inversion six-times. Samples were then centrifuged for 10mins at 13,000rpm and approximately 800µL plasmid DNA containing supernatant was transferred to QIAprep 2.0 spin columns to bind plasmid DNA. The supernatant was spun down for 1min at 13,000rpm and flow-through discarded. Columns were washed by the addition of 750µL Buffer PE and centrifuging for 1min at 13,000rpm. The flow-through was discarded, and residual wash buffer was removed by an additional centrifugation step. Columns were transferred to clean Eppendorf tubes, and DNA eluted by adding 50µL nuclease-free H₂O. Columns were left to stand for 1min at RT, followed by centrifuging columns for 1min at 13,000rpm. Plasmid DNA was stored at -20°C.

2.2.1.20 Plasmid amplification and isolation – maxi- and giga- preps

5mL LB media (with appropriate antibiotic) was inoculated with approximately 10µL bacteria glycerol stock or a single colony of transformed bacteria. Inoculations were grown for approximately 8h at 37°C, 200rpm to generate a starter culture. Starter culture was then used to inoculate 200mL LB media (with appropriate antibiotic) and grown overnight at 37°C, 200rpm. Plasmid DNA was then extracted using QIAfilter Plasmid Maxi Kit (Qiagen). Cultured bacteria cells were pelleted down by ultracentrifugation at 8000rpm for 15mins at 4°C.

Pelleted cells were resuspended in 10mL P1 Buffer, followed by the addition of 10mL P2 Buffer, with vigorous inversions six-times. Lysis reactions were performed at RT for 5mins, and then neutralised by the addition of 10mL pre-chilled P3 Buffer (3M potassium acetate, pH 5.5) with vigorous inversions six-times. Neutralised reactions were immediately transferred to the barrel of the QIAfilter Cartridge, and incubated up-right for 10mins at RT. The QIAGEN-tip 500 was then equilibrated using 10mL Buffer QBT (750mM NaCl, 50mM MOPS, pH 7.0, 15% isopropanol [v/v], 0.15% Triton™ X-100 [v/v]) until the solution flowed through completely. The cell lysate was filtered into the equilibrated QIAGEN-tip 500. The lysate was allowed to enter the resin to bind plasmid DNA by gravity flow, and the flow-through was discarded. The QIAGEN-tip 500 was washed twice with 30mL Buffer QC (750mM NaCl, 50mM MOPS, pH 7.0, 15% isopropanol), and flow-through discarded. Plasmid DNA was eluted using 15mL Buffer QF (1.25M NaCl, 50mM Tris-HCl, pH 8.5, 15% isopropanol [v/v]) and collected in a clean tube. Plasmid DNA was then precipitated by mixing in 10.5mL RT isopropanol and centrifuged immediately at 12,000rpm for 30mins, 4°C. The supernatant was decanted and DNA pellet washed with 5mL RT 70% ethanol. The washed pellet was further centrifuged at 12,000rpm for 10mins and supernatant decanted. The pellet was air-dried at RT for approximately 5mins, and DNA pellet finally dissolved in 500µL nuclease-free H₂O. Alternatively, for larger quantities of plasmid the QIAfilter Plasmid Giga Kit (Qiagen) was used, and was performed similarly to QIAfilter Plasmid Maxi Kit. However, starter cultures and overnight bacteria cultures were scaled-up five-fold; bacteria pellet was resuspended and lysed in 125mL of P1, P2, and P3 Buffers. Lysates were then transferred to a QIAfilter Mega-Giga Cartridge and left to incubate for 10mins, RT. This was followed by vacuum-filtering the lysate until most of the liquid had been pulled through. 50mL of Buffer FWB2 (1M potassium acetate, pH 5.0) was added to the remaining lysate and gently mixed, and then completely vacuum filtered. The QIAGEN-tip 1000 was equilibrated with 75mL QBT Buffer, and after filtered lysate had completely run through the QIAGEN-tip 10000, the QIAGEN-tip was washed with a total of 600mL of Buffer QC. Plasmid DNA was eluted with 100mL of Buffer QF into a clean tube, followed by the addition of 70mL RT isopropanol to precipitate the DNA. The mixture was then centrifuged immediately at 12,000rpm for 30mins, 4°C. The supernatant was carefully decanted and DNA pellet was washed with 10mL of 70% ethanol, followed by centrifugation at 12,000rpm for 10mins, 4°C. The supernatant was carefully decanted and the DNA pellet air-dried for approximately 10mins. The DNA pellet was then dissolved in 1mL of nuclease-free H₂O, and stored at -20°C.

Harvesting bacmid DNA, however, involved the use of the Qiagen Plasmid Maxi Kit. Transformed DH10Bac™ colony or glycerol stock was inoculated in 4mL LB medium containing 50µg/mL kanamycin, 7µg/mL gentamicin, and 10µg/mL tetracycline, and grown overnight at 37°C, 250rpm to generate a starter culture. The overnight culture was then used to inoculate 200mL of LB medium with appropriate antibiotics, and incubated overnight at 37°C, 250rpm. Bacmid was then harvested similarly as described for QIAfilter Plasmid Maxi Kit, however, neutralised lysis reactions were incubated on ice for 20mins. Lysates were not filtered, and use of the QIAGEN-tip 500 was omitted because the size of recombinant bacmid DNA is too large to bond efficiently to the QIAGEN-tip 500. Instead, the lysate was pelleted down at 16,000rpm for 15mins. The clarified supernatant was then mixed with 10.5mL RT isopropanol and bacmid DNA eluted as described QIAfilter Plasmid Maxi Kit.

2.2.1.21 Measuring DNA concentrations

DNA concentration of DNA fragments and plasmids were quantified using a Nanodrop™ 1000 spectrophotometer (Thermo Fisher Scientific) after blank measurement with 1µL of the corresponding dissolvent. Next, 1µL of undiluted plasmid or eluted DNA fragment was applied to the lower measurement pedestal and the sampling arm was then brought down to bring the upper measurement pedestal in contact with the loaded sample. Spectral measurement was then initiated to estimate DNA concentration. This was achieved by measuring the sample's absorbance at 260nm (A260), and assuming that 1 A260 unit corresponds to 50µg/mL of dsDNA. DNA purity was assessed by the A260/A280 ratio, where a ratio of approximately 1.8-2.0 was considered pure DNA, in that the sample was relatively free from protein and RNA contaminants.

2.2.1.22 Measuring RNA concentrations

RNA concentrations were measured using a Qubit™ 3.0 Fluorometer (Thermo Fisher Scientific) and Qubit™ RNA BR Assay Kit (Invitrogen). RNA-specific fluorescent dyes bind to RNA species within a given sample, and the fluorescence emitted can be quantified, and made relative to calibrated samples of known concentration. RNA sample was first diluted 1:100 in nuclease-free H₂O. The Qubit™ RNA BR Reagent (composition – propriety information) was diluted 1:200 in Qubit™ RNA BR Buffer (composition – propriety information) to make Qubit™ Working Solution. Sample preparation was then made in 0.5mL Qubit™ Assay Tubes (Invitrogen). 1µL of diluted RNA was mixed in 199µL of Working Solution and 10µL of standard was mixed in 190µL of Working Solution, each. Standards and samples were briefly vortexed

for 3secs, and incubated at RT for 2mins. Standards were then read by the Qubit™ 3.0 Fluorometer to calibrate measurements, and concentrations of RNA samples were quantified. Final concentrations of RNA were adjusted to take into account the 1:100 dilution factor.

2.2.1.23 Sanger sequencing

Sanger sequencing was performed by Source Bioscience to confirm cloning and characterise genomic sequences of CRISPR/Cas9-genome edited cell lines. The following were prepared: plasmid DNA was provided at 100ng/μL concentration, PCR product was provided as 10ng (per 100bp)/μL concentration, and, sequencing-designated oligos (Table 2.2) were provided at 3.2pmol/μL concentration. All samples and primers were submitted to Source Bioscience in tube format (5μL of sample DNA or primer per sequencing reaction) using SpeedREAD™ service. Sequencing results were received online and analysed using BioEdit Sequence Alignment editor.

2.2.1.24 Plasmid constructs

Plasmid constructs that were generated by cloning and used in this thesis are presented below. The cloning strategy employed for the construction of particular plasmids is also described.

CRISPR constructs

CRISPR constructs includes the GeneArt™ linearised, all-in-one CRISPR nuclease vector (Invitrogen), pCRISPR-YBX1sgRNA1-4, and pCRISPR-SFYBsgRNAsf1-3 constructs. The pCRISPR constructs were generated by direct cloning of 5nM sgRNA DNA designs (see section 2.2.1.1 for details on sgRNA designs) after annealing oligo pairs listed in Tables 2.3 and 2.4 together, as per section 2.2.1.7. This allowed for 3'-overhangs to be established (forward oligo: 5'-GTTTT-3', and reverse oligo: 5'-CGGTG-3'), which are compatible with the 3'-overhangs present in the GeneArt™ linearised, all-in-one CRISPR nuclease vector, and resulted in gRNA expression under U6 promoter control. CRISPR constructs also harboured the CMV-promoter driven SpCas9-2A-OFP expression cassette. CRISPR constructs were verified by sequencing using the U6.F(orward) oligo (Table 2.2).

AAV plasmids

Plasmids utilised for rAAV2 vector production in mammalian cell lines included pAAV2-hrGFP (Stratagene), pAAV2/2-RC (Stratagene), pHelper (Stratagene), pAAV2-MCS (Stratagene), and pAAV2-FLuc. Plasmid pAAV2-hrGFP transfer vector includes the *hrGFP* transgene under CMV

promoter control, and a β -globin intron, flanked by AAV2 *ITRs*. Plasmid pAAV2/2-RC harbours the AAV2 *rep* and *cap* genes required for rAAV2 vector genome replication and packaging. Plasmid pHelper harbours the Adenovirus type 5 helper functions required *in trans* for rAAV2 vector production: E2A, VA RNAs, and E4orf6. Plasmid pAAV2-MCS is an expression vector, of which multiple cloning site is under CMV promoter control and β -globin intron, and flanked by AAV2 *ITRs*. Plasmid pAAV2-FLuc was generated by sub-cloning of the Firefly luciferase coding sequence (*FLuc*) from pJET1.2-FLuc (a kind gift from Kam Zaki, University College of London) as a BamHI-HindIII fragment. Firstly, pJET1.2-FLuc was double restriction enzyme digested with BamHI (NEB) and NotI (NEB) followed by gel purification. Next a NotI-HindIII adapter molecule was generated by annealing 5'-GCTGCGGCCGAGTCGACCTGCAGAAGCTTGCT-3' and 5'-AGGCAAGCTTCTGCAGGTGACTGCGGCCGAGC-3' oligos together, and double restriction enzyme digested with NotI and HindIII (NEB), followed by PCR purification. The adapter molecule and purified BamHI-NotI *FLuc* fragment were then ligated using T4 DNA Ligase and then PCR purified. Finally the pAAV2-MCS expression vector was double digested with BamHI (Promega) and HindIII (Promega), followed by dephosphorylation of DNA ends using rSAP and gel purified. The BamHI-HindIII vector backbone and BamHI-HindIII *FLuc* fragment were ligated together with T4 DNA Ligase. Positive cloning was verified by analytical restriction enzyme digestion of pAAV2-FLuc with BamHI and HindIII.

Lentiviral vector plasmids

Plasmids used to transfect 293T cells for lentiviral vector production included pRRSIN.cPPT.PKG-GFP.WPRE (a gift from Didier Trono; Addgene plasmid# 12252), pRSV-REV (a gift from Didier Trono; Addgene plasmid# 12253), pMDLG/pRRE (a gift from Didier Trono; Addgene plasmid# 12251), pMD2.G (a gift from Didier Trono; Addgene plasmid# 12259). The pRRSIN.cPPT.PKG-GFP.WPRE is a third generation lentiviral transfer vector encoding *EGFP* under hPGK promoter control and WPRE terminator flanked by HIV1 5'-LTR (truncated) and self-inactivating 3'-LTR (Δ U3). Plasmid pRSV-REV is a third generation lentiviral packaging plasmid encoding *rev* under RSV promoter control, where Rev is required for efficient export of lentiviral transcripts to the cytoplasm. Plasmid pMDLG/pRRE is another third generation packaging plasmid encoding *gag* and *pol* under CMV promoter control and β -globin intron, and code for the structural proteins and viral enzymes, respectively, required for functionally transducing lentiviral vectors. Plasmid pMD2.G is the VGV-G envelope expressing plasmid required for lentiviral vector infectivity. Finally, pDUAL.mIL6-puro (a kind gift from Illaria Nisoli, University College London) is a third generation lentiviral transfer vector that encodes mouse

IL6 (*mIL6*) under SFFV promoter control and WPRE terminator, and puromycin resistance gene under ubiquitin promoter control, all flanked by HIV1 5'-LTR and self-inactivating 3'-LTR (Δ U3) from HIV1.

YB1 expressing construct and YB1 Δ 1-6 lentiviral transfer vectors

Exogenous YB1 (myc-tagged) expression was mediated by transfection of mammalian cell lines using pDESTmycYBX1 plasmid (a kind gift from gift from Thomas Tuschl; Addgene plasmid# 19878). Lentiviral vector constructs encoding YB1 full length and YB1 Δ 1- Δ 6 truncation mutants were generated: pDUAL.YBX1-puro, pDUAL.YBX1 Δ 1-puro, pDUAL.YBX1 Δ 2-puro, pDUAL.YBX1 Δ 3-puro, pDUAL.YBX1 Δ 4-puro, pDUAL.YBX1 Δ 5-puro, and pDUAL.YBX1 Δ 6-puro, respectively; and were derived from pDUAL.mIL6-puro backbone. These constructs were generated by High-Fidelity PCR amplification of the YB1 full length coding sequence and YBX1 Δ 1- Δ 6 mutants using pDESTmycYBX1 plasmid as template. Amplification of full length *YBX1* coding sequence used YBF1 and YBR1; amplification of *YB1 Δ 1* used YBF1 and YBR2; amplification of *YB1 Δ 2* used YBF1 and YBR3; amplification of *YB1 Δ 3* used YBF1 and YBR4; amplification of *YB1 Δ 4* used YBF2 and YBR3; amplification of *YB1 Δ 5* used YBF3 and YBR1; amplification of *YB1 Δ 6* used YBF2 and YBR1 (see Table 2.2). PCR amplicons included 5'-NotI and 3'-BamHI restriction enzyme sites, Kozak sequence, and stop codon where appropriate. Following which, gel purified PCR products were double restriction enzyme digested with NotI (NEB) and BamHI (NEB) to generate NotI-BamHI fragments. These were cloned into gel purified pDUAL.mIL6-puro backbone (after NotI and BamHI double enzyme restriction enzyme digestion and dephosphorylation of DNA ends), using T4 DNA Ligase. Positive transformants were verified by double restriction enzyme digestion with BamHI and NotI, and by sequencing using SFFV.F(oward) and WPRE.R(everse) oligos (Table 2.2).

***His-SfYBco* encoding constructs for Sf9 transfections**

Plasmids for transfection of insect cell lines included: pIExTM-1 expression vector (Novagen), pIExTM-1.eGFP, pIExTM-1.*His-SfYBco*, pre-linearised pTriExTM1.1 (a kind gift from Prof. Ian Jones, University of Reading), of which pTriExTM-1.*His-SFYBco* was derived from. The pIExTM-1 plasmid is an insect cell line-transfection compatible expression vector, which drives desired recombinant protein expression under the control of hr5 enhancer and the IE1 promoter, and IE1 terminator. Cloning into pIExTM-1 multiple cloning site provided an in-frame Histidine-tag (deca-His-tag) encoded sequence for expression of His-tagged recombinant protein where intended. The pTriExTM-1.1 vector enables the expression of desired recombinant protein

under the control of CMV or baculovirus-derived p10 promoters, and also provides an in-frame deca-His-tag encoded sequence for His-tagged recombinant protein expression after cloning. The pTriEx™-1.1 also harbours flanking Tn7 right and left transposon elements for directional transposition.

The pEx-1.eGFP construct was generated by digesting pEx™-1 with NcoI (NEB) restriction enzyme, and then sub-cloning of the *EGFP* coding sequence from pFBGR as an NcoI fragment. Firstly, the pFBGR was double restriction enzyme digested with HindIII (NEB) and AgeI (NEB) to remove a third, undesirable NcoI(3) restriction site. After gel purification, cohesive ends were blunted using Klenow Fragment and DNA ends phosphorylated. Blunted pFBGR was then PCR purified for self-ligation using T4 DNA ligase. The resultant plasmid (pFBGRΔNcoI(3)) was then restriction enzyme digested with NcoI for *EGFP* fragment dropout and isolated by gel purification. The pEx™-1 plasmid was also restriction enzyme digested with NcoI and gel purified to isolate the vector backbone. After PCR purification the *EGFP* fragment was ligated into the gel purified pEx™-1 vector backbone using T4 DNA Ligase to generate pEx-1.eGFP plasmid. Positive transformants were screened by colony PCR using pEx1-ie1prom.F(oward) and eGFP.colPCR.R(everse) oligos (Table 2.2).

Plasmid pEx-1.His-SFYBco vector construct was generated by cloning of codon optimised *SFYB* (*SFYBco*) gene-string for *S. frugiperda* expression. *SFYBco* gene-string also included 5'-BamHI and 3'-HindIII restriction sites, and was synthesised by Thermo Fisher Scientific, GeneArt (Table 2.10). Codon optimisation involved substituting wildtype *SFYB* cDNA sequences for improved codon usage for *S. frugiperda* expression, as well as improved codon quality distribution and optimised GC content compared to wildtype *SFYB* cDNA sequence (Table 2.10), and multiple sequence alignment (MSA) using NCBI global align tool with wildtype *SFYB* as reference ('Sbjct') co*SFYB* as 'Query' (Fig. 2.1). Therefore, 200ng of *SFYBco* gene string was double restriction enzyme digested with BamHI and HindIII and gel purified. Next, the pEx™-1 expression vector was double restriction enzyme digested with BamHI and HindIII, and DNA ends subjected to dephosphorylation. This was followed by gel purification to isolate the vector backbone. The BamHI-HindIII *SFYBco* fragment was cloned into the vector backbone using T4 DNA Ligase. This allowed for the *SFYBco* fragment to clone into the pEx™-1 vector backbone in-frame of the vector's start codon and immediate deca-His-tag encoded sequence. Positive transformants were screened by colony PCR using pEx1-ie1prom.F(oward) and

pIEx1-ie1term.R(verse) oligos (Table 2.2), and plasmids by analytical restriction enzyme digestion using BamHI and HindIII.

Finally, pTriEx™-1.His-SFYBco was generated by PCR amplification of the *SFYBco* sequence using Promega 2X Master Mix, SFYB-inf.F(oward) and SFYB-inf.R(verse) oligos (Table 2.2), and pIEx™-1.His-SFYBco template. The PCR product was then gel purified and cloned by In-Fusion® (Clontech) into the pre-linearised pTriE™x-1.1 cloning vector. Positive transformants were verified by analytical restriction enzyme digestions using EcoRI and NcoI, and Sanger sequencing using the T7.F(oward) oligo (Table 2.2).

Table 2.10 wildtype *SFYB* and *SFYBco* gene string sequence

cDNA	5'>3' sequence
wildtype	<p>ATGGCTGATACCGAAAAGGCGCCGAGCCGAGCCCAACAGCAGCAGCAACAAGAG CAACAACCCCCACAGCAACCGCAACAAGCTAAAGCGGCTAAACAAAAGCAGGTCATTG CTGAGAAAGTTTCGGGAACCGTCAAATGGTTTAATGTCAAGAGTGGATATGGTTTCAT CAACAGGAATGACACCAAGGAGGATGTGTTTGTGCATCAAACCTGCAATCGCCCGGAAC AACCTCGCAAGGCTGTGCGCTCGGTGGCGACGGGGAGGCGGTGGAGTTTGCCGTG GTTGCCGGGGAGAAAGGCTATGAAGCAGCCGGAGTAACTGGTCCCGGTGGTGAGCCG GTAAAGGGCTCGCCCTATGCAGCTGACAAACGCCGCGGCTTCCATCGCCAATATTACCC CCGTCAAGGTGGCGGACGTGGCGGGGAAGGCGCTCCACGTAGAGGTGGAATGGGAC</p>
<i>SFYB</i>	<p>GTCGTGGGCCCCGACCAACCAGGGGGGTGCACAGGGGGATGAAGGTCAGGAGGGA GGCGGAGCACCACAGCGCAGCTACTTCCGCCGCAACTCCGTGGTGGACGCCGTG GTGGCGGTCCAGGGCCCATGAATCGCGGAGGATACCGTCGCGCTCGTCCACGCAACTT CCAACCGGGCCAAGGACAGGGCCAGCCCCAAGCTGGTGGTCAGCCTAACCAAGCACCA CGTCAAATGGTCAGGAGGCAGAGGCCCCAGCCACCGCTGCGTCGCCAACGCAACAGC AACAGGCCAAGCCGAAATCTACCACCAAGCCTGCTGGTACTACCATTGAGACCACTACC AATGAGAG CCAGGCCTAA</p>
<i>SFYBco</i>	<p>GCGC<u>GGATCCC</u>GCTGACACTGAAAAGGCTCCTCAGCCTCAGCCACAGCAGCAACAACA GCAAGAGCAGCAGCCTCCTCAGCAGCCCCAACAAGCTAAGGCTGCTAAGCAAAAAGCAA GTGATCGCTGAGAAGGTGTCCGGCACCGTGAAGTGGTTCAACGTGAAGTCCGGTTACG GTTTCATCAACCGCAACGACACCAAGAAGATGTGTTTCGTCCACCAGACCGCTATCGCT CGTAAACAACCCTCGCAAGGCTGTGCGTTCTGTGCGGAGATGGCGAGGCTGTGGAATTCCG CTGTGGTGGCTGGCGAGAAGGGTTACGAAGCTGCTGGTGTACTGGTCCCGGTGGCGA</p>

ACCTGTGAAGGGTTCTCCTTACGCTGCTGACAAGCGTCGTGGTTTCCACCGTCAGTACT
ACCCTCGTCAAGGTGGTGGTCTGGTGGCGAAGGTGCTCCTAGAAGAGGTGGAATGG
GTCGTCGTGGTCCTCCTACCAATCAAGGTGGCGCTCAAGGCGACGAGGGTCAAGAAGG
TGGCGGCGCTCCTCCACAGAGGTCTTACTTCCGTCGTAACCTCCGCGGTGGTCTAGAG
GTGGTGGACCCGGTCTATGAACCGTGGTGGTTATCGTCGTGCTCGTCCCCGTAATTC
CAGCCTGGTCAAGGACAGGGACAGCCCCAAGCTGGTGGACAGCCTAATCAGGCTCCTC
GTCAGAACGGTCAAGAGGCCGAAGCTCCTGCTACCGCTGCTTCTCCTACTCAGCAGCAG
CAGGCTAAGCCCAAGTCCACCCTAAGCCCGCTGGCACCACCATCGAGACTACCACCA
CGAGTCCCAGGCTTAAAGCTTGCGC

Bold underlined sequence, BamHI restriction site; **blue** bold sequence, wildtype *SFYB* or *SFYBco* coding sequences; double underlined sequence, HindIII restriction site.

NW Score	Identities	Gaps	Strand
765	648/825(79%)	0/825(0%)	Plus/Plus
Query 1	ATGGCTGATACCGAAAAGGCGCCGACGCCGACGCCAACAGCAGCAGCAACAAGAGCAA		60
Sbjct 1	ATGGCTGACTGAAAAGGCTCCTCAGCCTCAGCCACAGCAGCAACAAGAGAGCAG		60
Query 61	CAACCCACAGCAACCGCAACAGCTAAAGCGGCTAAACAAAAGCAGGTCATTGCTGAG		120
Sbjct 61	CAGCCTCCTCAGCAGCCCCAACAAAGCTAAGGCTGCTAAGCAAAAGCAAGTGATCGCTGAG		120
Query 121	AAAGTTTCGGGAACCGTCAAATGGTTTAAATGTCAAGAGTGGATATGGTTTCATCAACAGG		180
Sbjct 121	AAGGTGTCCGGCACCGTGAAGTGGTTCAACGTGAAGTCCGGTTACGGTTTCATCAACCGC		180
Query 181	AATGACACCAAGGAGGATGTGTTTGTGCATCAAAGTCAATCGCCCGGAACAACCCCTCGC		240
Sbjct 181	AACGACACCAAGGAGGATGTGTTTGTGCATCAAAGTCAATCGCCCGGAACAACCCCTCGC		240
Query 241	AAGGCTGTGCGCTCGGTCGGCGACGGGGAGGCGGTGGAGTTTGCCTGGTTGCCGGGGAG		300
Sbjct 241	AAGGCTGTGCGTTCTGTCCGGAGATGGCGAGGCTGTGGAATTCGCTGTGGTGGCTGGCGAG		300
Query 301	AAAGGCTATGAAGCAGCCGGAGTAACTGGTCCCGGTGGTGGAGCGGTAAAGGGCTCGCCC		360
Sbjct 301	AAGGGTTACGAAGCTGCTGGTGTACTGGTCCCGGTGGCGAACCTGTGAAGGGTCTCCT		360
Query 361	TATGAGCTGACAAACGCCGGGCTTCCATCGCCAAATATACCCCGTCAAGGTGGCGGA		420
Sbjct 361	TACGCTGCTGACAAGCGTCTGGTTCACCCGTCAGTACTACCCCTCGTCAAGGTGGTGGT		420
Query 421	CGTGGCGGGGAAGGCGCTCCAGTAGAGGTGGAAATGGGACGTCGTGGGCCCCGACCAAC		480
Sbjct 421	CGTGGTGGCGAAGGTGCTCCTAGAAAGAGGTGGAAATGGGTGCTGTGGTCTCCTACCAAT		480
Query 481	CAGGGGGGTGCACAGGGGGATGAAGGTCAAGGAGGGAGGGCGGAGCACCACCACAGCGCAGC		540
Sbjct 481	CAAGGTGGCGCTCAAGGCGACGAGGGTCAAGAAAGGTGGCGGCGCTCCTCCACAGAGGTCT		540
Query 541	TACTTCCGCCCAACTTCCGTTGGTGGACGCCGTTGGTGGCGGTCAGGGCCCATGAATCGC		600
Sbjct 541	TACTTCCGTCGTAACCTCCGCGGTGGTGGTAGAGGTGGTGGACCCGGTCTATGAACCGT		600
Query 601	GGAGGATACCGTCGCGCTCGTCCACGCAACTTCCAACCGGGCCAAGGACAGGGCCAGCCC		660
Sbjct 601	GGTGGTTATCGTCGTGCTCGTCCCGTAATTTCCAGCCTGGTCAAGGACAGGGACAGCCC		660
Query 661	CAAGCTGGTGGTCAAGCCTAACCAAGCACCACGTCAAAATGGTCAGGAGGCAGAGGCCCA		720
Sbjct 661	CAAGCTGGTGGACAGCCTAATCAGGCTCCTCGTCAGAACGGTCAAGAGGCCGAAGCTCCT		720
Query 721	GCCACCGCTGCGTCGCCAACGCAACAGCAACAGGCCAAGCCGAAATCTACCACCAAGCCT		780
Sbjct 721	GCTACCGCTGCTTCTCCTACTCAGCAGCAGCAGGCTAAGCCCAAGTCCACCCTAAGCCC		780
Query 781	GCTGGTACTACCATTTGAGACCACTACCAATGAGAGCCAGGCCATAA 825		
Sbjct 781	GCTGGCACCAACATCGAGACTACCACCAACGAGTCCAGGCTTAA 825		

Figure 2.1 BLAST/MSA analysis of wildtype vs codon optimised *SFYB* cDNA sequences. Wildtype ('Sbjct') and codon optimised *SFYB* ('Query') cDNA sequences were inputted into NCBI Global align tool, and cDNA sequence similarity assessed by MSA. BLAST analysis indicated that codon optimisation of *SFYB* by GeneArt™ introduced some nucleotide substitutions, resulting in 79% sequence identity to wildtype *SFYB*.

pFastBac plasmids

pFastBac plasmids used to transform DH10Bac™ *E. coli* included pFBGR (a kind gift from Robert Kotin; Addgene plasmid# 65422), pSR657 (a kind gift from Robert Kotin; Addgene plasmid# 65214), and pSR660 (a kind gift from Robert Kotin; Addgene plasmid# 65216). The pFBGR functioned as the shuttle vector for the *EGFP* transgene under the control of CMV enhancer and promoter or p10 promoter, flanked by AAV2 *ITRs*. Plasmid pSR657 functioned as the shuttle vector harbouring AAV2 *rep* and *cap* ORFs, under promoter control of polh and p10, respectively. Finally, pSR660 functioned as a shuttle vector encoding AAV2 *rep* and AAV8 *cap* ORFs, under promoter control of polh and p10, respectively. All pFastBac plasmids harboured the Tn7L and Tn7R transposon elements required for transposition of desired rAAV-specific sequences into bacmid DNA to generate recombinant bacmid DNA. Positive transposition was verified by growing transformed DH10Bac™ cells on LB agar containing Blue-White Select™ Reagent.

2.2.2 In silico methods

2.2.2.1 Predictive alternative splice site activation and alternative splicing motifs

Alternative splice site activation was assessed using Human Splicing Finder (HSF) v3.0 (<http://www.umd.be/HSF3/HSF.html>; Desmet *et al.*, 2009), and simulating the mutations that were detected by sequencing. Alternative splicing predictions were calculated against wildtype *YBX1* splicing patterns and motifs (ENST00000321358) as reference and using the HSF Matrices prediction algorithm (Desmet *et al.*, 2009). Potential alternative transcripts identified by HSF were input into protein translator tool (<https://web.expasy.org/translate/>) to conceptually translate and determine encoded amino acid (aa) sequence, and compared to wildtype YB1 aa sequence (NP_004550.2) as reference. Multiple sequence analysis was also performed in which the extrapolated aa sequence for each potential alternative transcript was input into NCBI global align tool (as 'Query'), against wildtype aa sequence (NP_004550.2) as reference ('Sbjct').

2.2.2.2 CRISPR/Cas9 off-target screening

In order to identify potential off-target sites for potential CRISPR/Cas9-mediated disruption with the sgRNA designs used, the sgRNA DNA designs were separately input into NCBI BLAST online tool using the Reference Genomic Sequences (refseq_genomic) database (Altschul *et al.*, 1997), and restricting organism selection to *H. sapiens* (taxid: 9606). The top five off-target sites for each gRNA DNA design were selected for further analysis.

2.2.2.3 Putative identification of Y-Box protein homologues

The selected cDNA sequences were acquired from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). BLAST assessments were carried out using NCBI's Transcriptome Shotgun Assembly (TSA) database (Zhang *et al.*, 2000) with organism refinement limited to *Spodoptera* (taxid: 7106) or *Trichoplusia ni* (taxid: 7111). Selected candidate cDNA sequences were then input into a protein translator tool (<https://web.expasy.org/translate/>) to conceptually translate candidate cDNA(s) in all forward and reverse reading frames. Candidate frame readouts from start methionine to stop codon was then input into NCBI's Global Alignment tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=GlobalAln&BLAST_PROGRAMS=blastn), which uses the Needleman-Wunsch alignment method (Needleman and Wunsch, 1970). Candidate sequences were assigned as subject (Sbjct), against human YB1 (NP_004550.2) or BYB (NP_001036897.1) sequences as references (Query), to reverse BLAST and verify aa sequence homology.

2.2.2.4 Multiple sequence alignments (MSA) and phylogeny

Selected aa sequences were conceptually translated and then inputted into the PRALINE MSA programme (<http://www.ibi.vu.nl/programs/pralinewww/help.php>) (Simossis *et al.*, 2005) to construct and analyse MSA. The evolutionary history was inferred by phylogeny using the bootstrapped (Felsenstein, 1985) UMGMA method (Sneath and Sokal, 1973) of MSA data. Phylogenetic analysis of aligned proteins was conducted by MEGA X (Kumar *et al.*, 2018). The evolutionary distances were computed using the Poisson correction method (Zuckerandl and Pauling, 1965) and are in the units of the number of aa substitutions per site. All positions containing gaps and missing data were eliminated.

2.2.2.5 Predictive intrinsic disorder

Intrinsic disorder was predicted by inputting aa sequences into DISOclust (<http://www.reading.ac.uk/bioinf/IntFOLD/>). Prediction of disordered residues was calculated based on probability scoring from comparing multiple models on a per-residue basis, and thus scores of <0.5 was considered structured and ≥ 0.5 was considered disordered (McGuffin, 2008). Additionally, intrinsic disorder was also predicted using PSIPRED and DISOPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) after inputting of selected aa sequences. PSIPRED and DISOPRED predicts intrinsic disorder, secondary structures, and/or protein binding sites by

examining patterns of sequence conservation on intrinsically disordered regions (Jones and Cozzetto, 2015).

2.2.2.6 Predictive hydrophobicity

Amino acid sequences were inputted into ExPASy ProtScale programme (<https://web.expasy.org/protscale/>) for hydrophobicity scales using the Kyte and Doolittle classifications and scoring criteria (Kyte and Doolittle, 1982), and a window size of 9. Residues were considered hydrophobic with scores >0.

2.2.2.7 Comparative 3D-structure protein modelling

3D-models were generated using IntFOLD server (<http://www.reading.ac.uk/bioinf/IntFOLD/>), by inputting selected aa sequences and running template-based 3D-modelling with model quality assessments (McGuffin *et al.*, 2015). Of the resultant structures modelled, the top scoring model's .pdb file were then exported and rendered in the Jmol: an open-source Java viewer for chemical structures in 3D (<http://www.jmol.org/>) for interactive visualization. This was followed by use of TM-align (<https://zhanglab.ccmb.med.umich.edu/TM-align/>) to compare and computationally assesses 3D-model structure similarities along superimposed structures (Zhang and Skolnick, 2005), using the template modelling (TM)-score between template structures provided by IntFOLD server and between Y-Box protein homologues. The TM-score was scored between 0-1, with 1 implying a perfect match, and TM-score of 0.5<1 indicated that the modelled proteins are in about the same fold.

2.2.2.8 Annotating and mapping *SFYB* gene sequence

S. frugiperda putative *SFYB* cDNA sequence (GCTM01011706.1) was inputted into the NCBI BLAST tool with organism refinement set to *Spodoptera* (taxid: 7106) using the Whole-genome Shotgun Contigs database to identify the assembled contig encoding the *SFYB* gene. The *BYB* (YB1 orthologue) gene was identified using NCBI's Genome Data Viewer (<https://www.ncbi.nlm.nih.gov/genome/gdv/>), and exons and introns were mapped using the genomic scaffold (NW_004582015.1) and available transcriptomics (GCF_000151625.1). *S. litura* Y-Box protein homologue gene was mapped using NCBI's Genome Data Viewer (NC_036217.1) and available transcriptomics (LOC111360813). Using both Y-Box protein homologue gene annotations as genomic scaffolds for *SFYB*'s gene, its exons and introns were mapped for the assembled contig accession OE0A01010394.1 based on sequence homology.

2.2.3 Cell culture, transfections and infections of mammalian and insect cell lines

2.2.3.1 Cell culturing conditions

293T and cell-derivatives were cultured in complete Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich) supplemented with 10% heat-inactivated foetal calf serum (FCS), 2mM L-glutamine (Sigma-Aldrich), and 50U/mL penicillin and 50µg/mL streptomycin (Sigma-Aldrich). Cells were cultured at 37°C and 5% CO₂. Passaging cells involved washing monolayer cells with sterile 1X phosphate buffered saline (PBS; 137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄, pH 7.4; sterilised by autoclaving), followed by brief treatment with Trypsin-EDTA solution (Sigma-Aldrich) diluted 1:10 in sterile 1X PBS. Cells were allowed to detach from the plate with gentle rocking. Trypsin-EDTA was inactivated by the addition of equal volume of complete DMEM. Cell clumps were dissociated by pipetting cells up-and-down. Sf9, Sf-9ET, and Ao38 cells were cultured in Sf-900™ II serum-free media (Sf-900™ II SFM; Gibco), and 50U/mL penicillin and 50µg/mL streptomycin (complete Sf-900™ II SFM), supplemented with 2% heat-inactivated FCS for Ao38 and Sf-9ET cells. Insect cells were cultured either as adherent cells or in suspension. Adherent monolayers were grown using suitable sandwich boxes with hydration supplied by facial tissue dampened with sterile H₂O, which was frequently replaced. Sandwich boxes were first washed with 70% ethanol and every three days for extended static cell culturing. Cells were incubated at 27°C using a refrigerated incubator. Suspension cells were instead maintained at 27°C at a constant 135rpm rotation. Adherent Sf9 cells would be passaged by disrupting the adherent monolayer in flasks by forcefully hitting the base of the flask against a hard surface. Cells were then well resuspended by pipetting to dissociate cell clumps.

With respects to seeding plates or flasks for a controlled density, well resuspended cells were first counted by haemocytometer. The number of cells within each central 5x5 grid was counted, and an average count was calculated. In the case of insect cell lines, cells were first mixed with an equal volume of Trypan Blue Solution, 0.4% (Gibco). Cells were then transferred to haemocytometer chambers. The average number of cells counted (*N*) corresponded to *N* x10⁴ cells/mL. The cell density calculated for insect cell lines were multiplied by two to account for the dilution factor. The % of viable insect cells was determined by 1 – (number of stained cells/number of total cells) x100. Sf9, Sf-9ET and Ao38 cells were passaged and used as log-phase growing cells exhibiting >95% viability. Ultimately, a known number of cells were transferred to new flasks or plates as indicated, in a suitable volume of their corresponding

complete media. This was followed by gentle rocking to distribute the cells homogenously across the surface of plate or flasks.

2.2.3.2 Freezing and thawing cells

Mammalian cell lines were trypsinised with Trypsin-EDTA solution and cell density/mL counted as described in section 2.2.3.1. Cell suspension was then transferred to a falcon tube and cells pelleted down at 1000rpm for 5mins. Supernatant was removed and the cell pellet was resuspended in freezing solution (composed of 90% FCS and 10% DMSO). 1mL suspensions corresponding to approximately $3-5 \times 10^6$ cells were aliquoted into freezing vials. Cryopreservation of insect cell lines, however, involved counting cells as described in section 2.2.3.1, followed by pelleting cells for 5mins at 1500rpm. After which the conditioned media was aspirated and mixed into equal volume of fresh Sf-900™ II SFM, and then mixed with DMSO to constitute 7.5% of the final volume. Cells were transferred as 1.5mL aliquots, corresponding to $1.5-2 \times 10^7$ cells, into freezing vials. Vials were slowly frozen at -80°C overnight, and then transferred into liquid nitrogen vapour phase tanks for indefinite storage at -150°C .

Thawing either cells involved removing frozen cells from liquid nitrogen storage, and thawing cells briefly at 37°C . This was followed by resuspending thawed cells in their corresponding complete media (pre-warmed). Mammalian cells were then pelleted for 5mins, 1000rpm to remove DMSO. The supernatant was removed, and replaced with suitable volume of pre-warmed complete media. Well resuspended cells were then transferred to suitable flasks or plates, and then rocked gently to homogenously distribute cells.

2.2.3.3 Single-cell cloning by limiting dilution

Mammalian or insect cell lines were harvested and resuspended well in their corresponding complete culture media. The cell density was quantitated by haemocytometer and with Trypan Blue exclusion where indicated. The cell suspension was serially diluted in their corresponding culture media with or without particular supplements (as indicated), to achieve the required cell densities. 100µL of diluted cell suspension was then aliquoted into 96-well plates. An additional 50µL of the corresponding culture media was finally added to each well to increase the final volume to 150µL/well. Cells were then cultured at 37°C and 5% CO_2 or 27°C (for mammalian and insect cell lines, respectively) until clonal cell populations were identified. Clonal cell populations were then grown to confluency for expansion.

2.2.3.4 Measuring cell growth

Log phase-growing cells were collected and cell densities calculated by haemocytometer with Trypan Blue exclusion. Cells were then passaged into 125mL Erlenmeyer flasks at a seeding density of 3×10^5 cells/mL per flask, in a total 30mL complete Sf-900™ II SFM. Cells were cultured at 27°C, 135rpm. Cell densities and viable cell counts were then counted by haemocytometer and Trypan Blue exclusion for up to 14 days, with live cell densities/mL and % of viable cells recorded at approximately 24h intervals.

2.2.3.5 Lentiviral vector production

293T (5×10^6) cells were seeded into 15cm dishes in 25mL complete DMEM per plate, and cultured at 37°C and 5% CO₂ to achieve approximately 70% confluency the next day. Cells were then transfected using the Calcium Phosphate precipitation method. On the day of transfection, on a per plate basis, 5µg pMD.2G, 5µg pRSV.REV, 10µg pMDLg/pRRE, and 20µg of transfer vector was mixed to 0.9mL final volume in sterile ultra-pure H₂O. This was followed by the addition of 100µL of 2.5M CaCl₂ (0.45µm sterile filtered) and 1mL of 2X HBS (1.5M NaHPO₄, 50mM HEPES, 280mM NaCl, pH 7.05, 0.22µm sterile filtered) dropwise. Mixes were incubated at RT for 5mins, then the precipitation was inactivated by the addition of 2mL pre-warmed complete DMEM (37°C). 293T cells were then treated with chloroquine (25mM stock; prepared by diluting chloroquine diphosphate [Sigma-Aldrich] in ultra-pure H₂O and 0.22µm filter sterilised) at 25µM final concentration. Cells were transfected with 4mL of transfection mix added dropwise with swirling to mix. Transfections were returned to culture at 37°C and 5% CO₂. 72h post-transfection, culture media was harvested and centrifuged at 4000rpm for 10mins. The supernatant was then filtered through a 0.45µm filter, and transferred to ultra-clear Beckman ultracentrifuge tubes (ultracentrifuge tubes were initially rinsed with 70% ethanol and washed with sterile 1X PBS, followed by air-drying). Tubes were balanced with the addition of sterile 1X PBS for ultracentrifugation of lentiviral vectors for 90mins, 28,000rpm at 4°C using a Beckman Coulter Optima XL ultracentrifuge and SW32 Ti rotor. After which, the supernatant was carefully decanted and excess liquid drained by inverting tubes onto tissue. Lentiviral vector pellets were then soaked in 350µL opti-MEM™ (approximately 100-fold concentration), and pellet resuspended by pipetting up-and-down sixty-times. Resuspended pellets were incubated on ice for 1h, after which, the resuspended pellet was pipetted up-and-down again sixty-times. Lentiviral vector suspension was then transferred to screw-capped Eppendorf tubes and stored at -80°C.

2.2.3.6 Lentiviral vector titration

293T cells were seeded in 500 μ L at 6×10^4 cells/well, in complete DMEM into 24-well plates. Cells were then cultured overnight at 37°C and 5% CO₂. On the day of infection, three wells of 293T cells were trypsinised and counted by haemocytometer to establish the approximate number of cells at the time of infection. Next, lentiviral vector stock was then serially diluted in pre-warmed (37°C) opti-MEM™ with polybrene (Sigma-Aldrich) at 8 μ g/mL final concentration to achieve 1:5, 1:25, 1:100, 1:500, 1:2500, and 1:5000 dilutions. Complete DMEM was then aspirated from 293T cells, and 200 μ L of titrated lentiGFP was added carefully in triplicate for each titration. 200 μ L of opti-MEM™ with polybrene was added to 293T cells in three separate wells to function as negative infection control. Plates were returned to incubate at 37°C and 5% CO₂ for 6h. The media was then replaced with 500 μ L of complete DMEM. 72h post-infection, cells were harvested and fixed in final 1% formaldehyde, and then analysed by flow cytometry using negative infection control cells to gate against GFP-positive populations. Samples expressing 1-30% GFP were considered for calculating lentiGFP vector titres. Titres were calculated as transducing units (TU)/mL by $TU/mL = (F \times C_n) / (Vol \times DF)$. F corresponds to the frequency of GFP-positive cells determined by flow cytometry (between 1-30%), C_n corresponds to the number of cells on the day of infection, Vol corresponds to the inoculum volume of lentiviral vector, and DF refers to the dilution factor. Final TU/mL was calculated as the overall average.

2.2.3.7 Lentiviral vector transduction

Mammalian cells were seeded into 24-well plates in 500 μ L at 1×10^5 cells/well of complete DMEM, or into 6-well plates at 3.5×10^5 cells/well in 2mL of complete DMEM, and cultured overnight at 37°C and 5% CO₂ for approximately 70% confluency the next day. On day of infections, for single well infections, 50 μ L, 60 μ L or 300 μ L of lentiviral vector, or volumes corresponding to the desired MOI were diluted into pre-warmed (37°C) opti-MEM™ with polybrene (final 8 μ g/mL) to a final volume of 200 μ L or 1mL, for 24-well plates and 6-well plates, respectively. Complete DMEM was aspirated and diluted lentiviral vector was added to cells. Cells were incubated at 37°C and 5% CO₂ for approximately 6h, after which the inoculum was replaced with 500 μ L or 2mL of complete DMEM for 24-well plates and 6-well plates, respectively. Cells were then further incubated at 37°C and 5% CO₂ for 72h and then expanded. Transduced cells were then selected with puromycin (Invivogen) by replacing cultured media with complete DMEM supplemented with puromycin (final 3 μ g/mL

concentration). Transduced cell lines were maintained in complete DMEM supplemented with puromycin for several passages, and then maintained in complete DMEM.

2.2.3.8 Transient plasmid transfection of mammalian cell lines – Lipofectamine™ 2000

Cells were seeded into 6-well plates at 3.5×10^5 cells/well in 2mL complete DMEM, and cultured overnight at 37°C and 5% CO₂ for approximately 70-80% confluency the next day. Transfections were performed following manufacturer's instructions for Lipofectamine™ 2000 Reagent (Invitrogen). On a per well basis, transfection mixes were prepared by diluting plasmid DNA of the required quantities into opti-MEM™ for a final volume of 150µL. 6µL of Lipofectamine™ 2000 Reagent was separately diluted in opti-MEM™ to a final volume of 150µL. The dilutions were mixed together and incubated at RT for 5mins, and 300µL of transfection mix was then added dropwise to cultured cells. Cells were cultured at 37°C and 5% CO₂ for 72h or as indicated.

2.2.3.9 Transient plasmid transfection of mammalian cell lines – FuGENE® HD

Cells were seeded into 6-well plates at 3.5×10^5 cells/well in 2mL complete DMEM, respectively, and cultured overnight at 37°C and 5% CO₂ for approximately 70-80% confluency the next day. Transfections were performed following the manufacturer's instructions for FuGENE® HD Reagent (Promega). Firstly, on a per well basis, plasmid DNA at the required quantities was diluted to 300µL with pre-warmed opti-MEM™ (37°C) and briefly vortexed for 5secs. A 3:1 FuGENE® HD Transfection Reagent:DNA ratio was used, and added to the diluted plasmid with a brief vortex of 10secs to mix. The transfection mix was then incubated for 15mins at RT, and approximately 300µL of transfection mix was added dropwise to cultured cells. Transfections of cells in 24-well plates were scaled down accordingly. Transfected cells were cultured for 72h, or as indicated, at 37°C and 5% CO₂.

2.2.3.10 Transient plasmid transfection of mammalian cell lines – rAAV2 vector production

Cells were seeded into 15cm dishes at 5×10^6 cells/plate in 25mL complete DMEM or in triple flasks at 2×10^7 cells/flask in approximately 80mL complete DMEM. Cells were cultured overnight at 37°C and 5% CO₂ for approximately 70-80% confluency on day of transfection. Transfection for rAAV vector production was performed using Calcium Phosphate precipitation method with chloroquine (Sigma-Aldrich) or PEI_{max} (Polysciences). On a per plate basis transfections using Calcium Phosphate precipitation were prepared by diluting plasmid DNA (7.5µg pAAV2-hrGFP, 7.5µg pHelper, and 22.5µg pAAV2/2-RC) in sterile H₂O to make a final

volume of 0.9mL. 100µL of sterile 2.5M CaCl₂ was added, followed by 1mL of 2X HBS dropwise. Mixes were incubated at RT for 5mins, then the precipitation was inactivated by the addition of 2mL pre-warmed complete DMEM (37°C). At which point chloroquine was added to cultured cells for a final concentration of 25µM. 4mL of the Calcium Phosphate transfection mix was then added to the cultured cells dropwise with swirling to mix. Cells were incubated for 72h (or as indicated) at 37°C and 5% CO₂. Cells transfected for rAAV vector production or with AAV plasmids using Lipofectamine™ 2000 or FuGENE® HD Reagents were performed using scaled-down plasmid quantities described for Calcium Phosphate precipitation. Additionally, transfection mixes followed the corresponding protocols described in sections 2.2.3.8 and 2.2.3.9, respectively.

Cells transfected for rAAV vector production using PEI_{max} per triple-flasks was performed by diluting plasmids (13.8µg pAAV2-hrGFP, 13.8µg pAAV2/2-RC, and 41.4µg of pHelper) to a final volume of 5mL in pre-warmed opti-MEM™ (37°C). PEI_{max} (1mg/mL stock prepared in ultra-pure H₂O, pH 7.0 and 0.22µm sterile filtered) was warmed to RT and vortexed for 20secs prior to use. Following which, 241.5µL of PEI_{max} (3.5:1 PEI_{max} to plasmid DNA ratio) was diluted to 5mL final volume of pre-warmed opti-MEM™ (37°C). Diluted parts were mixed together and vortexed for 20secs, followed by 20mins RT incubation. Approximately 10mL of transfection mix was added to 80mL fresh complete DMEM, and used to replace the culture media per triple-flask. Transfection of cells in 15cm dishes using PEI_{max} were scaled-down accordingly, and added dropwise to cultured cells with swirling to mix, instead. Cells were incubated for 72h at 37°C and 5% CO₂.

2.2.3.11 Transient plasmid transfections of Sf9 cell line

Sf9 cells were transfected as per Cellfectin™ II (Invitrogen) or Lipofectamine™ CRISPRmax™ Cas9 Transfection Reagent (Invitrogen) manufacturer's instructions. For Cellfectin™ II, log-phase growing Sf9 cells were seeded into 6-well plates in 2mL complete Sf-900™ II SFM at 8.5x10⁵ cells/well. Cells were allowed to adhere as a monolayer for 1h at RT. After which, the media was replaced with 2.5mL Plating Medium (Grace's Insect Medium [Gibco] supplemented with 1.5% FCS, 7.5U/mL penicillin, and 7.5µg/mL streptomycin). On a per well basis transfection mixes were prepared by diluting plasmid DNA at the required quantity in 100µL Grace's Insect Medium and diluting 8µL Cellfectin™ II reagent in 100µL Grace's Insect Medium. Dilutions were then mixed both together by pipetting and incubated at RT for 30mins. The transfection mix (approximately 210µL) was then added dropwise to cells in

Plating Medium and transfected for 5h at 27°C. The media was then replaced with 2mL complete Sf-900™ II SFM, and transfected cells were returned to 27°C for 72h. Transfection of Sf9 cells using Cellfectin™ II Reagent in 24-well plates was scaled down accordingly. Alternatively, transfections of suspension Sf9 cells using Cellfectin™ II reagent, log-phase growing Sf9 cells were seeded at 1×10^7 cells in 8mL complete Sf-900™ II SFM in Erlenmeyer flasks. Transfection mixes using Cellfectin™ II was prepared by scaling up transfection mixes according to cell number from 6-well format and optimum input plasmid quantity. Transfected cells were returned to 27°C at 125rpm culture for 72h.

Whereas, log-phase growing Sf9 cells transfected using Lipofectamine™ CRISPRmax™ Cas9 Transfection Reagent were seeded into 24-well plates in 500µL complete Sf-900™ II SFM at 2×10^5 cells/well, and allowed to adhere for 1h at RT. On a per well basis, plasmid DNA at the desired quantity and 1µL of Lipofectamine™ Cas9 Plus™ Reagent were then diluted in 25µL opti-MEM™, Sf-900™ II SFM, or Grace's Insect Medium. 1.5µL of CRISPRmax™ Reagent was separately diluted in 25µL of the corresponding SFM. The respective diluted mixes were then mixed together and vortexed for 10secs and incubated at RT for 15mins. Approximately 50µL of transfection mix was then added dropwise and cells incubated at 27°C for 72h. Transfections of Sf9 cell lines for recombinant baculovirus (P0) stocks is described in section 2.2.3.13.

2.2.3.12 Transfection of Sf9 cells with recombinant Cas9 and sgRNA

Log-phase growing Sf9 cells were seeded into 96-well plate in 100µL of complete Sf-900™ II SFM at 3×10^4 cells/well. Cells were allowed to adhere as monolayer cells for 1h at RT. Transfection mixes were prepared as per GeneArt™ Platinum™ Cas9 Nuclease (Invitrogen) instructions, with some modifications. On a per well basis, 50ng of recombinant Cas9 with 0.1µL of Lipofectamine™ Cas9 Plus™ Reagent, and 12.5ng of sgRNA was diluted in 4µL of Grace's Insect Medium. 0.15µL of CRISPRmax™ Reagent was separately diluted in 4µL of Grace's Insect Medium. Both dilutions were mixed together and vortexed for 10secs, followed by incubation at RT for 15mins. For 10:1 input sgRNA plus recombinant Cas9 to CRISPRmax™ Transfection Reagent ratio, the input quantities of sgRNA, recombinant Cas9 and Lipofectamine™ Cas9 Plus™ Reagent was scaled up 10-fold. Transfection mixes were then added to Sf9 cells and cells incubated at 27°C for 72h. Non-transfected Sf9 cells was prepared in parallel as negative controls.

2.2.3.13 Recombinant baculovirus production

Recombinant baculoviruses were generated using either the Bac-to-Bac® (Invitrogen) or *flashBAC*™ (OET) Baculovirus Expression Systems. For *flashBAC*™ Baculovirus Expression System, log-phase growing Sf9 cells were seeded into 6-well plates in 1mL complete Sf-900™ II SFM at a cell density of 2×10^6 cells/well. Sf9 cells were allowed to adhere as monolayer cells for 1h at RT to achieve approximately 80% confluency. This was followed by preparation of the co-transfection mix: 500ng of appropriate plasmid (harbouring Tn7 transposon elements) was mixed with 100ng of *flashBAC*™ DNA (OET) and 1.2µL baculoFECTIN Reagent (OET), prepared in Sf-900™ II SFM and incubated for 15mins at RT. Cells were then co-transfected with *flashBAC*™ DNA:plasmid mix overnight at 27°C. The next day, 1mL of complete Sf-900™ II SFM was added, and cells incubated at 27°C for a further 4 days. The medium containing baculovirus (P0 stock) was then collected, and used to propagate to P1 and P2 stocks.

Alternatively, for Bac-to-Bac® Baculovirus Expression System, log-phase growing Sf9 cells were seeded into 6-well plates in 1mL complete Sf-900™ II SFM at a density of 8.5×10^5 cells/well, and allowed to adhere for 1h at RT. After which, the media was replaced with 2.5mL Plating Medium. Transfections were performed as described in section 2.2.3.11 using 500ng of bacmid DNA and 8µL of Cellfectin™ II Reagent. However, transfected cells were incubated at 27°C for 5 days. Recombinant baculovirus (P0 stock) was harvested from the cultured media, and transferred to clean falcon tubes with FCS added to approximately 2% of harvested media volume. Baculovirus stocks were stored at +4°C.

2.2.3.14 Amplification of baculovirus stocks

An accelerated P1 baculovirus stock was amplified by seeding a T75 flask of Sf9 cells (4.5×10^6 log-phase growing cells) in 10mL complete Sf-900™ II SFM and allowing cells to adhere for 1h at RT. Following which, the harvested P0 baculovirus stock generated using the *flashBAC*™ Baculovirus Expression System was added directly to cultured Sf9 cells and infected for 3-5 days for P1 baculovirus stock expansion. After which the culture media was harvested. Otherwise, general amplification of the P0 recombinant baculovirus to P1/P2 involved seeding log-phase growing Sf9 cells into Erlenmeyer flasks in 30mL complete Sf-900™ II SFM at 2×10^6 cell/mL density. Cells were then infected with recombinant baculovirus (P0 or P1 stock to generate P1 and P2 stocks, respectively) at MOI of 0.05 prepared in complete Sf-900™ II SFM, made to 10% volume of the culture volume. Infected cells were incubated at 27°C at 135rpm for 7 days, after which the culture media was harvested. Harvested cells were pelleted at

1500rpm for 5mins. The recombinant baculovirus-containing supernatant was transferred to new falcon tubes with FCS added to approximately 2% of the harvest volume. P1 and P2 stocks were stored at +4°C.

2.2.3.15 Plaque assays

Recombinant baculovirus stocks were quantified by plaque assays on monolayer Sf9 cells. Log-phase growing Sf9 cells were seeded into 6-well plates in 2mL complete Sf-900™ II SFM at 1×10^6 cells/well, and allowed to adhere for 1h at RT. Baculovirus stock was then serially diluted 10-fold in complete Sf-900™ II SFM to 10^{-8} dilution. After cell attachment the media was replaced with 1mL inoculums of 10-fold serially diluted baculovirus (10^{-4} to 10^{-8} dilutions), in triplicate per dilution factor, for 1h at RT. The inoculum was then removed and cells overlaid with Plaquing Media (final 1% low melting temperature agarose [Lonza] in Sf-900™ II SFM [1.3X, Gibco], warmed to 40°C). The Plaquing Media was allowed to briefly set and then cells were incubated for 10 days at 27°C. After which, plates were stained with 0.5mL of Neutral Red Solution (Sigma-Aldrich), prepared as 1mg/mL solution in ultra-pure H₂O, for 2h at RT. The Neutral Red Solution was then drained away by inverting plates on tissue, to allow the enumeration of plaques. Plates that exhibited between 2-40 plaques were considered only. Approximate recombinant baculovirus titres were calculated by: average number of plaques x corresponding dilution factor, and then multiplying by 1/inoculum volume (1mL). Final average plaque forming units (pfu)/mL was then calculated.

2.2.3.16 Titration of recombinant baculovirus by endpoint dilution

Log-phase growing Sf-9ET cells were seeded into 96-well plates at 7.5×10^4 cells/well in 100µL complete Sf-900™ II SFM plus 5% FCS (100µL per well), and allowed to adhere for 1h at RT. Baculovirus stocks were then serially diluted in complete Sf-900™ II SFM with 5% FCS from 10^{-1} to 10^{-8} , and cells were inoculated with 20µL of diluted baculovirus/well in replicates of eight for each dilution factor. Non-infected controls included inoculating Sf-9ET cells with 20µL of Sf-900™ II SFM plus 5% FCS only, in triplicate. Infected cells were cultured at 27°C for 72h, after which the number of wells exhibiting foci of infection (positive foci of infection were defined as a cluster of Sf-9ET cells that exhibited green fluorescence by fluorescent microscopy) (Hopkins and Esposito, 2009). Virus titres were determined by calculating the median tissue culture infectious dose (TCID₅₀) following the Reed-Müench method (Reed and Muench, 1938). Here, the proportionate distance (PD) was calculated by $PD = (\text{fraction of wells next above } 50\% \text{ infection rate}) - 50\% / (\text{fraction of wells next above } 50\% \text{ infection rate}) - (\text{fraction}$

of wells next below 50% infection rate). TCID₅₀ was then calculated by subtracting PD from the negative logarithm of the dilution next above 50% infection rate. TCID₅₀/mL was calculated by multiplying TCID₅₀ by inoculum volume (mL). Finally, final titres were calculated by multiplying TCID₅₀/mL by 0.69 for approximate recombinant baculovirus titres as pfu/mL.

2.2.3.17 Infection of insect cells with recombinant baculovirus

Log-phase growing Sf9 or Ao38 cells were seeded in 6-well plates at 2×10^6 cells/well in final 2mL of complete Sf-900™ II SFM (supplemented with 5% FCS for Ao38 cells), and allowed to adhere for 1h at RT for approximately 80% confluent monolayers. Where appropriate, the seeding density and culture volume was scaled-down for smaller tissue culture plates. Alternatively, log-phase growing Sf9 cells were seeded at 2×10^6 cells/mL density into Erlenmeyer flasks. Recombinant baculovirus(es) (corresponding to MOI of 3) was diluted in complete Sf-900™ II SFM to a final inoculum volume corresponding to 10% of the culture volume. For rAAV vector production, BacITRGFP and BacAAV2.2 or BacAAV2.8 were co-infected at MOI 3 each. Diluted recombinant baculovirus(es) was then added dropwise to seeded Sf9 or Ao38 cells and infected for the indicated time, and cultured at 27°C (and 135rpm for infected Sf9 in suspension culture).

2.2.3.18 Crude AAV vector harvest

Transfected mammalian cells or infected insect cells for rAAV vector production were harvested into their culture media. Cells in 15cm dishes were scraped off into the culture media and cells in triple flasks were trypsinised and pooled. Cells were then pelleted down at 1500rpm for 15mins. Cells were thrice washed with sterile 1X PBS. The final PBS wash was then completely aspirated carefully, and cells resuspended in 1mL TD Buffer (140mM NaCl, 5mM KCl, 0.71mM K₂HPO₄, 25mM Tris-HCl [pH 6.8], 6.4mM MgCl₂, 0.22µm filter sterilised). The volume of TD Buffer used to resuspend transfected mammalian cells lines was scaled-up accordingly for transfections performed in triple-flasks. Alternatively 5mL of TD Buffer was used to resuspend insect cells infected for rAAV vector production. Resuspended cells were immediately placed on dry-ice and subjected to five lots of freeze-thaws in dry ice and 37°C water bath until cells were completely frozen or thawed, respectively. Cell debris was then pelleted at 4500rpm for 20mins and supernatant collected and transferred to clean Eppendorf tubes and -20°C, or subjected to purification by an iodixanol gradient.

2.2.3.19 Purification of rAAV vectors

Different iodixanol gradients were prepared (15%, 25%, 40%, and 60% concentrations) using OptiPrep™ Gradient Medium (Sigma-Aldrich). OptiPrep™ was diluted to 15% using ultra-pure H₂O with NaCl to a final 1M concentration; and to 25% in final 1X TD Buffer (5X: 250Mm Tris-HCl [Ph 8.5], 750mM NaCl) and 0.00125% Phenol Red Solution (Sigma-Aldrich) using ultra-pure H₂O; and to 40% using in final 1X TD Buffer using ultra-pure H₂O. Finally, the 60% gradient was prepared by mixing in Phenol Red Solution to a final 0.00125%. The gradient was assembled slowly and stepwise by adding 1.55mL of the 60% gradient, followed by 1.55mL of the 40% gradient, then 1.88mL of the 25% gradient, and finally 2.8mL of 15% gradient to Beckman ultracentrifuge tubes. 5mL of crude rAAV vector lysate was then treated with 2000U of Benzonase® Nuclease (Sigma-Aldrich) for 1h at 37°C. 3mL of crude rAAV vector lysate was then slowly added per tube, and tubes balanced. The rAAV vector was then separated from empty capsids and recombinant baculovirus by ultracentrifugation for 3h at 40,000rpm and 18°C using a Beckman Coulter Optima XL ultracentrifuge and SW40 Ti rotor. Therefore, rAAV vector was collected from between the 40-60% layers using a 5mL syringe and 20-gauge needle. The collected fraction was pooled for corresponding rAAV vector samples, and concentrated using 100kDa cut-off Vivaspin 6 Concentrator (Sartorius). This involved the buffer exchange of the OptiPrep™ with 1X PBS. Therefore, the pooled rAAV vector sample was mixed with 6mL of 1X PBS and applied to the Vivaspin 6 Concentrator, and centrifuged at 1000rpm for 5mins intervals. Between each centrifugation interval the diluted rAAV vector sample was mixed by pipetting to homogenise the rAAV vector. When approximately 0.5-1mL of sample was left, the flow-through was discarded and the Vivaspin 6 Concentrator was topped up with 1X PBS to 6mL. Centrifugation at 5mins intervals, 1000rpm and intermittent homogenising the rAAV vector and buffer exchange repeated for a total of three buffer exchanges with 1X PBS. Finally, approximately 300µL of concentrated rAAV vector was collected, and volumes normalised between corresponding batches to 300µL using 1X PBS. Purified rAAV vector was stored at -20°C.

2.2.3.20 Gene transfer assays

293T cells were seeded into 24-well plates at 1×10^5 cells/well in 500µL complete DMEM, and cultured overnight at 37°C, 5% CO₂. Crude rAAV vector samples were harvested by five-time freeze-thaws as described in 2.2.3.18. Collected crude rAAV was then diluted in pre-warmed opti-MEM™ (37°C) at 1/5, 1/25, and 1/100 dilutions, and then treated at 60°C for 20mins to heat-inactivate baculovirus(es). Crude lysate (neat) derived from five-time freeze-thaws of

BacITRGFP+BacMers-M-infected Sf9 cells was also heat-inactivated for 20mins at 60°C as a control for complete baculovirus heat-inactivation. Crude lysates derived from five-time freeze-thaw of BacMers-M-infected Sf9 cells functioned as mock infection control of 293T cells. Therefore, 293T cells were infected with 50µL crude lysates at the indicated dilutions. 293T cells were also infected with 50µL non-heat-inactivated crude lysate derived from Sf9 infected with BacITRGFP+BacMers-M control to help establish the efficiency of baculovirus heat-inactivation. Infections were carried out for 48h at 37°C and 5% CO₂. Cells were then harvested and analysed by flow cytometry against mock infected control for negative gating of GFP-positive populations.

2.2.3.21 Lactate dehydrogenase (LDH) cytotoxicity assays

Mammalian cells were seeded in 15cm dishes at 5x10⁶ cells in 25mL complete DMEM per dish or 6-well plates at 3.5x10⁵ cells/well in 2mL complete DMEM. Alternatively, Sf9 cells were seeded into 6-well plates at 2x10⁶ cells/well in 2mL complete Sf-900™ II SFM and allowed to form a monolayer for 1h at RT. Cells were transfected with AAV-plasmids for rAAV vector production as indicated (described in sections 2.2.3.8-10), or infected with baculoviruses for rAAV vector production (described in section 2.2.3.17). Alternatively, cells were treated with the intended reagents – including transfection mixes with the omission of plasmids, or chloroquine only, or non-treated. Final input volumes of each treatment were all normalised by each cell line's respective complete media or opti-MEM™. In all cases complete media (with 5% FCS where appropriate) was added to separate 6-well plates alone and cultured in parallel for consideration of LDH activity attributable by FCS and for normalisation of LDH cytotoxicity activities.

LDH activities were measured using the Pierce™ LDH Cytotoxicity Assay (Thermo Fisher Scientific). At the indicated time points 150µL of culture media was collected and transferred to 96-well plates in triplicate for each condition tested and each replicate. Cells from baseline reference controls (non-treated or treatments with transfection mixes with the omission of plasmids, where indicated) were then harvested in their culture media. Control cells were then pipetted up-and-down to dissociate cell clumps. 150µL of these resuspended cells were transferred to 96-well plates in triplicate for each replicate. 150µL of FCS-containing (where relevant) media was also transferred to 96-well plates in triplicate for background LDH activity assessment and normalisation. Plates were then centrifuged at 2000rpm for 3mins. 15µL of 10X Lysis Buffer (composition – proprietary information) was added to wells containing

harvested cells, and functioned as maximum LDH activity controls (100% lysis). Plates were gently tapped to mix and then incubated at 37°C and 5% CO₂ for 45mins. 50µL of each sample was transferred to new 96-well plates, and mixed with 50µL of Reaction Mixture, which was composed of lyophilized Substrate Mix (composition – proprietary information) dissolved in 11.4mL ultra-pure H₂O and mixed with 0.6mL of Assay Buffer (composition – proprietary information). Plates were incubated for 30mins at RT, and protected from light in order for LDH to catalyse the conversion of lactate to pyruvate via the reduction of NAD⁺ to NADH. A colorimetric reaction ensued that utilises the NADH to reduce a tetrazolium salt to a red formazan product. Reactions were terminated by the addition of 50µL LDH Stop Solution (composition – proprietary information). Bubbles were disrupted by centrifugation at 2000rpm for 3mins, and plates read by spectrophotometer at 490nm and 680nm. LDH activity was calculated by subtracting 680nm absorbance readings from 490nm for each sample, to eliminate background absorbance. Following which, the LDH cytotoxicity was calculated as a % of max lysis by: %cytotoxicity = (LDH activity – LDH activity present in the complete media) / (maximum LDH activity – LDH activity present in the complete media) x100. Average %cytotoxicity was then calculated.

2.2.4 Harvesting genomic DNA and protein

2.2.4.1 Genomic DNA extraction

Genomic DNA was harvested using the Wizard[®] Genomic DNA Purification Kit (Promega). Cells were routinely harvested, and approximately 3-5x10⁶ cells were centrifuged at 13,000rpm for 10secs. Supernatant was then removed and washed in 200µL PBS. Cells were pelleted at 13,000rpm for 10secs and supernatant removed. 600µL of Nuclei Lysis Solution (composition – proprietary information) was added and cell pellet resuspended until no cell clumps remained. 17.5µL of 20mg/mL Proteinase K (Qiagen) was added, and mixtures incubated at 55°C for 3h with intermittent vortexing. Mixtures were allowed to cool to RT for 5mins, and 3µL of RNase A Solution (4mg/mL RNase A in DNA Rehydration Solution: 10mM Tris-HCl [pH 7.4], 1mM EDTA [pH 8.0]) was then added to lysates and incubated for 30mins at 37°C. Lysates were cooled to RT for 5mins, after which 200µL of Protein Precipitation Solution (composition – proprietary information) was added and immediately vortexed for 20secs. Samples were incubated on ice for 5mins, and precipitate then centrifuged for 4mins at 13,000rpm. The supernatant containing genomic DNA was transferred to clean Eppendorf tubes, and centrifuged at 13,000rpm for 4mins. The clarified supernatant was transferred to new Eppendorf tubes and 600µL of RT isopropanol was added. Mixtures were mixed by inversions

until white, thread-like strands of genomic DNA formed. The precipitated genomic DNA was centrifuged at 13,000rpm at RT for 1min, and supernatant decanted. The pelleted genomic DNA was washed with RT 70% ethanol by gentle inversions, and then centrifuged at 13,000rpm for 1min. The ethanol was completely aspirated and genomic DNA pellet left to air-dry for 5mins. Following which, 100µL of DNA Rehydration Solution was added to rehydrate and dissolve the genomic DNA for 1h at 65°C. Genomic DNA was finally stored at -20°C.

2.2.4.2 Whole cell protein extraction

Cells were either lysed in RIPA Lysis Buffer (150mM NaCl, 1% Triton™ X-100 [v/v], 0.5% sodium deoxycholate, 0.1% SDS [w/v], 50mM Tris-HCl [pH 8.0]) or Lysis Buffer A (25mM HEPES [pH7.5], 100mM KCl, 12.5mM MgCl₂, 1mM DTT, 0.1% IGEPAL-CA630 [v/v], 10% glycerol [v/v]), supplemented with HALT™ EDTA-free Protease Inhibitor Cocktail (Thermo Fisher Scientific) diluted 1:100. Cells were first washed with ice-cold 1X PBS twice, and then scrapped into a suitable volume (200µL for a single well of a 6-well plate or 2mL for 15cm dishes) of pre-chilled RIPA Lysis Buffer or Lysis Buffer A, and kept on ice for approximately 15mins. Cell debris was then pelleted by centrifugation at 13,000rpm for 10mins, 4°C to collect the supernatant (lysate) into clean Eppendorf tubes. Lysates were stored at -20°C.

Additionally, where indicated Sf9 cells were harvested 72h post-transfection and pelleted at 1500rpm for 5mins. Cells were then lysed in 1mL of 0.1% Triton™ X-100 in PBS with cOmplete™ Protease Inhibitor Cocktail, EDTA-free (Roche), and homogenised with ten-strokes using a Dounce homogeniser on ice. Lysis reactions were then incubated on ice for 20mins followed by a brief vortex. Cell debris was pelleted at 13,000rpm for 15mins and supernatant transferred to clean Eppendorf tubes and kept on ice for subsequent pulldown using Dynabeads™ His-Tag Isolation and Pulldown (see section 2.2.5.2). Alternatively, 96h post-infection of Sf9 cells with recombinant baculovirus, approximately 2L of cells were harvested by centrifugation at 4500rpm for 15mins, 4°C. Cells were then resuspended in 20mL of 0.1% Triton™ X-100 in PBS, with cOmplete™ Protease Inhibitor Cocktail, EDTA-free, and then sonicated for 10mins, at 20secs on/20secs off pulse-rate, on ice. This was followed by centrifugation at 15,000rpm, 30mins at 4°C. The clarified supernatant was collected into a clean tube and kept on ice for subsequent purification by affinity chromatography (see section 2.2.5.5).

2.2.4.3 Cytoplasmic and membrane fractionation

Non-infected and baculovirus-infected Sf9 cells were harvested 96h post-infection, and cell densities counted by haemocytometer. 4×10^6 cells were pelleted down at 1500rpm for 5mins, 4°C, and resuspended in 300µL ice cold 0.1% Triton™ X-100 in PBS. Cells were lysed for 30mins on ice with intermittent vortexing every 5mins. Cell debris (membrane fraction) was pelleted down at 13,000rpm for 20mins, 4°C to collect the supernatant (cytoplasmic fraction). Following which, the cell debris was washed twice with 0.1% Triton™ X-100 in PBS and spun down at 13,000rpm, 4°C for 1min per wash. The supernatant was completely aspirated and the pellets were each resuspended in 60µL 5X Laemlli Buffer (250mM Tris-HCl [pH 6.8], 10% SDS [w/v], 25% glycerol [v/v], 0.02% Bromophenol Blue, 5% β-mercaptoethanol). Therefore, normalised loads were ensured for each fraction between samples. The membrane fraction was immediately boiled at 95°C for 10mins, and both fractions stored at -20°C. The cytoplasmic fraction was examined for nuclei/DNA contamination by mixing lysates with 6X DNA Loading Dye and subjecting samples to 1% TAE agarose gel electrophoresis.

2.2.5 Quantification and detection of proteins

2.2.5.1 Quantification of protein concentration

Protein concentrations were generally measured following the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) instructions. BSA standards were prepared in the corresponding lysis buffer used to lyse cultured cells, and involved diluting the BSA stock (2mg/mL) to achieve 2, 1.5, 1, 0.75, 0.5, 0.25, 0.125, and 0.025mg/mL standards. 25µL of standard and cell lysate sample was transferred to 96-well plates in duplicate, including 25µL of the corresponding lysis buffer to function as a blank standard. 200µL of BCA Working Reagent was added per well. Plates were mixed by gentle tapping, and protected from light and incubated at 37°C for 30mins. Plates were allowed to cool to RT and absorbance at 562nm was measured using a spectrophotometer. After blank correction, standards were used to generate a standard curve, from which the concentration of protein in cell lysate samples was determined. Alternatively, purified His-tagged recombinant protein concentration was measured by preparing known quantities of BSA standards (200, 400, 600, and 800ng), and loading 2, 4, 6, and 8µL of purified recombinant protein onto a 10% Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel. After SDS-PAGE, the gel was stained in Coomassie® Blue Solution then destained using Destain Solution (see section 2.2.5.6). Destained gels were then imaged and exported as .TIFF files, and analysed as 8-bit images by densitometry using ImageJ. A standard

curve was generated from the known quantities of BSA standards, from which the concentration of recombinant protein was calculated.

2.2.5.2 Recombinant His-tagged protein pulldown

Transfected or baculovirus-infected Sf9 cell lysates for recombinant His-tagged protein expression was purified by Dynabeads™ His-Tag Isolation and Pulldown (Invitrogen). 10µL of Dynabeads™ was first washed by resuspending beads in 1mL of 0.1% Triton™ X-100 in PBS with cOmplete™ Protease Inhibitor Cocktail, EDTA-free, followed by aspiration of the supernatant using a DynaMag™-2 Magnet for 2mins (magnetic rack; Thermo Fisher Scientific). Dynabeads™ were then mixed with 1mL of lysate for 30mins on ice. Beads were then washed thrice with 0.1% Triton™ X-100 in PBS with cOmplete™ Protease Inhibitor Cocktail, EDTA-free, and a magnetic rack for 2mins. His-tagged recombinant protein was then eluted in 100µL high Imidazole Elution Buffer (500mM Imidazole, 20mM NaHPO₄, 500mM NaCl) for 5mins, on ice, and supernatant transferred to clean Eppendorf tubes and stored at -20°C.

2.2.5.3 SDS-PAGE

Generally, 20µg of protein from whole cell lysates in RIPA Lysis Buffer or Lysis Buffer A was used. Whereas, 20µL of either cytoplasmic or nuclear fractions (corresponding to 2.6x10⁵ cells) was run; or the indicated volumes of protein samples was used, for subsequent SDS-PAGE. 5X Laemmli Buffer was added to 1X final concentration in sample lysates and mixed. Protein samples were denatured by boiling at 95°C for 10mins, and subsequently resolved on 10% SDS-PAGE mini-gels or pre-cast 4-12% NuPAGE™ Bis-Tris gels (Thermo Fisher Scientific). 10% SDS-PAGE mini-gels were prepared by adjusting the volume of 30% acrylamide (acrylamide:bis-acrylamide 29:1 solution; Bio-rad) in a solution made to final concentrations of 0.375M Tris-HCl (pH 8.8), 2.5µL/mL 10% SDS, 3.3µL/mL 10% APS, and 1.7µL/mL TEMED, in ultra-pure H₂O. Stacking gels were prepared by adjusting 30% acrylamide to 4% in a solution made to final concentrations of 0.125M Tris-HCl (pH 6.8), 1.7µL/mL 10% SDS, 3.3µL/mL 10% APS, and 1.7µL/mL TEMED, in ultra-pure H₂O. 10% SDS-PAGE mini-gels were set up in a Mini-PROTEAN® Tetra Cell System (Bio-rad), and immersed in 1X SDS-PAGE Running Buffer (10X: 30.3g Trizma-base, 144.1g glycine, 1% SDS, made to 1L ultra-pure H₂O). Pre-cast NuPAGE™ Bis-Tris gels were set up in XCell SureLock™ Mini-Cell (Invitrogen) electrophoresis system, and gels immersed with 1X MES SDS Running Buffer (20X: 1M MES, 1M Tris-HCl [pH 7.3], 2% SDS, 20mM EDTA). Boiled protein samples and 4µL of pre-stained protein marker as indicated – were loaded onto immersed 10% SDS-PAGE gels and run at 80V initially until sample and

Bromophenol Blue dye-front reached the separating gel, after which the voltage was increased to 150V until sample and Bromophenol Blue dye-front reached the bottom. 4-12% NuPAGE™ Bis-Tris gels were electrophoresed at constant 200V until the Bromophenol Blue dye-front reached the bottom.

2.2.5.4 Western blotting

Proteins were resolved by SDS-PAGE as described in section 2.2.5.3, after which proteins were transferred from SDS-PAGE gels to a Hybond® ECL™ nitrocellulose membrane (0.45µm pore size) using a Bio-Rad Mini Trans-Blot® Cell transfer system in pre-chilled 1X Transfer Buffer (10X: 30.3g Trizma-base, 144.1g glycine, made to 1L in ultra-pure H₂O) with 20% methanol (v/v). Protein transfer was performed at 100mA/blot for 3h on ice. Membranes were then blocked for 1h at RT with gentle agitation (70rpm) using 5% skimmed-dried milk (w/v) in PBS-T (0.1% Tween™ 20 [v/v] in 1X PBS). Primary antibodies were prepared in PBS-T with 5% skimmed-dried milk (w/v) at the required dilutions overnight at 4°C using a roller shaker. The following day, membranes were washed with PBS-T for 10mins, thrice with agitation (70rpm). This was followed by incubation in secondary HRP-conjugated αMouse (Sigma-Aldrich), αRabbit (Sigma-Aldrich), or αGoat (Dako) antibodies at the required dilution in PBS-T with 5% skimmed-dried milk (w/v) for 2h at RT, with agitation at 70rpm. Membranes were washed in PBS-T for 10mins thrice with agitation (70rpm). Blotted proteins were then detected by chemiluminescence with ECL™ Western Blotting Detection Reagent (Amersham), which was prepared according to the manufacturer's instructions by mixing ECL reagents 1 and 2 (1:1), and applying 1mL per blot. After 1min incubation with the ECL™ Western Blotting Detection Reagent mix, membranes were exposed on X-ray Hyperfilms (Amersham) for visualisation of proteins using an automated X-ray film processor.

2.2.5.5 Dot Blot

The indicated protein samples were blotted onto 1cm² squares of Hybond® ECL™ nitrocellulose membranes (0.45µm pore size) at the indicated quantities, and blots dried for 15mins at RT. Blots were blocked in PBS-T plus 5% skimmed-dried milk (w/v) for 1h at RT, with gentle agitation (70rpm). Blots were incubated with primary antibody (α6His-HRP; 1:5000) or antiserum at the indicated dilution, diluted in PBS-T plus 5% skimmed dried milk (w/v), for 1h at RT, 70rpm. This was followed by washes with PBS-T for 10mins, thrice with agitation (70rpm), followed by incubation in secondary HRP-conjugated αRabbit antibody at 1:2000 dilution in PBS-T with 5% skimmed-dried milk (w/v) for 2h at RT, with agitation at 70rpm.

Membranes were finally washed in PBS-T for 10mins, thrice with agitation (70rpm). Blotted proteins were then detected by chemiluminescence with ECL™ Western Blotting Detection Reagent, which was prepared according to the manufacturer's instructions by mixing ECL reagents 1 and 2 (1:1), and applying 100µL per square of blotted membrane. After 1min incubation with the ECL™ Western Blotting Detection Reagent mix, membranes were exposed on X-ray Hyperfilms for visualisation of proteins using an automated X-ray film processor.

2.2.5.6 Coomassie® Blue staining

10% SDS-PAGE gels were prepared and loaded with protein samples boiled at 95°C for 10mins. Protein samples were then resolved until the run-off the Laemlli loading dye-front. Following which, the 10% SDS-PAGE gel was stained in Coomassie® Blue Solution (0.2% Coomassie® Brilliant Blue R250 [w/v], 60% methanol [v/v], 8% acetic acid [v/v]) for 2h with gentle agitation. This was followed by destaining gels of background Coomassie® Blue stain using Destain Solution (25% ethanol [v/v], 8% acetic acid [v/v]) until background gel staining disappeared and protein bands were visible.

2.2.5.7 Silver staining

Silver gel staining of SDS-PAGE gels were performed following Peirce™ Silver Stain Kit (Thermo Fisher Scientific) instructions. After protein samples were subjected to SDS-PAGE, the gel was submerged and washed with ultra-pure H₂O (>18 megohm/cm resistance) for 5mins, twice. Gels were then fixed in 30% ethanol:10% acetic acid solution (prepared using ultra-pure H₂O) for 15mins, twice. This was followed by washing fixed gels in 10% ethanol solution (prepared with ultra-pure H₂O) for 5mins, twice. The gel was then incubated in Sensitizer Working Solution for 1min, followed by two washes at 1min each with ultra-pure H₂O. The Sensitizer Working Solution was initially prepared by mixing 1 part Silver Stain Sensitizer (composition – propriety information) with 500 parts ultra-pure H₂O. The washed gel was then incubated in Stain Working Solution, prepared by mixing 1 part Silver Stain Enhancer (composition – propriety information) with 50 parts Silver Stain (composition – propriety information) for 30mins. The gel was then washed in ultra-pure H₂O, twice, for 20secs each wash. The Developer Working Solution, prepared by mixing 1 part Silver Stain Enhancer with 50 parts Silver Stain Developer (composition – propriety information), was added and gels incubated until silver stained bands appeared. After which, the Developer Working Solution was quickly removed and replaced with Stop Solution (5% acetic acid [v/v] prepared in ultra-pure H₂O).

Gels were washed in Stop Solution for 1min, after which the Stop Solution was replaced and incubated for a further 10mins, and imaged.

2.2.5.8 Affinity chromatography

Clarified supernatant containing recombinant His-tagged protein was filtered through a 2µm glass fibre AP20 filter (Merck), and then subjected to immobilised metal affinity chromatography using a BioLogic LP System (Bio-rad) and a 5mL HisTrap HP column (GE Healthcare). Clarified and filtered supernatant containing recombinant His-tagged protein was then applied to a HisTrap HP 5mL column at 1mL/min flow rate in low Imidazole Binding Buffer (20mM Imidazole, 20mM NaHPO₄, 500mM NaCl) to allow the His-tagged recombinant protein to bind to the immobilised resin of Nickel Sepharose. This was followed by elution of the His-tagged purified recombinant protein from the immobilised Nickel Sepharose resin using high Imidazole Elution Buffer at 2.5mL/min flow rate. An in-run spectrophotometer was used to analyse the flow-through of total proteins and the elution of target His-tagged recombinant protein in 5mL elution fractions.

2.2.5.9 Flow cytometry

Mammalian or insect cell lines that were infected or transfected with GFP expressing vector or stained for intracellular protein were harvested (trypsinised for mammalian cell lines) and pelleted down at 1500rpm for 5mins. Following which, cells were washed in 1X PBS, and resuspended in 1% formaldehyde solution. Non and/or mock -transfected or -infected cell lines were prepared and fixed in parallel as negative controls for fluorescence, and used to gate fluorescence-positive cell populations. Fluorescence sorting was performed on a BD FACSCanto II flow cytometer (BD Sciences), and in-run data was acquired by FACSDiva Software (BD Sciences). Flow cytometry data was analysed on FlowJo 10.2 Software (FlowJo LLC).

2.2.5.10 Fluorescent microscopy

Cells were examined by phase contrast and fluorescence microscopy (using the FITC or Cy3 fluorescence cubes where appropriate) using an IX51 Inverted Microscope (Olympus) with a 10X objective lens. Images were captured with an XM10 Monochrome Camera (Olympus) attached to the inverted microscope and cellSens Imaging Software (Olympus).

2.2.5.11 Laser scanning confocal microscopy

Cells were seeded into plates containing circular, glass coverslips, thickness 1.5, 0.13mm at the required density. At the indicated points cells adhered to coverslips were washed with 1X PBS once, then fixed in 4% formaldehyde solution for 30mins at RT. Cells were then permeabilised with 0.1% Triton™ X-100 in 1X PBS for 10mins at RT, and washed again three times. Fixed cells were blocked in Blocking Buffer (5% FCS [v/v], 5mg/mL BSA [w/v], in 1X PBS) for 1h at RT. Cells were then incubated with primary antibodies at the indicated dilutions (table 2.7) in Blocking Buffer for 1h at RT, with gentle agitation at 70rpm, or overnight at 4°C. After primary antibody incubation, cells were washed with Blocking Buffer five times, followed by incubation with Alexa-Fluor- or DyLight- conjugated secondary antibodies at the required dilutions (Table 2.8) for 1h at RT and gentle agitation at 70rpm. Staining controls included cells treated with only primary or secondary antibody. Cells were finally washed with Blocking Buffer six-times, and coverslips were mounted onto glass slides with DAPI mounting medium with Fluoroshield™ (Sigma-Aldrich). Images were taken on a Leica SPX upright laser scanning confocal microscope with a 63X objective lens, and images analysed by Leica Application Suite X software and ImageJ.

2.2.5.12 Orthogonal cross-sections

To rule out z-plane discrepancy of stained cells for fluorescent signals detected by confocal laser scanning microscopy, orthogonal cross-sections of merged z-stacks was acquired using Leica Application Suite X software. Orthogonal cross-sections images included the x-, y- and z-planes and were exported as .TIFF images.

2.2.5.13 Line-profiling

Merged confocal images were analysed by Leica Application Suite X software by drawing separate lines (ROI) through three individual cells, and included the nucleolar compartment. The nucleolar compartment was distinguished by high resolution DAPI-staining. Grey intensities were measured along the length of each ROI (µm) for target protein fluorescence and DAPI staining, and plotted against the position along the ROI.

2.2.5.14 Quantification of corrected total cell fluorescence (CTCF)

Confocal images were exported as .TIFF files and converted to 8-bit images. Using ImageJ, the total cell fluorescent intensities were calculated by drawing ROIs around individual cells (20 ROIs per cell line, except of 5 ROIs for negative staining control – secondary antibody staining

only) and recording individual grey intensities. Background grey intensity was measured by taking the average of 5 ROIs from areas outside of the cell. Therefore, individual cell grey intensities were background corrected per measured cell line, and a final average grey intensity, corresponding to the CTCF, was then calculated.

2.2.5.15 Quantification of nuclear:cytoplasmic ratios (N/C)

Confocal images were exported as .TFF files and converted to 8-bit images. Mean fluorescent grey intensities in the nuclear and cytoplasmic regions were quantified from 30-40 individual cells per cell line, by drawing circular region(s) of interest (ROI) within the respective compartments using ImageJ. DAPI fluorescence distinguished the nuclear compartment of stained cells. Background grey intensity was measured by taking the average of 5 ROIs outside of the cells. Therefore, background-corrected mean grey intensities for the nuclear and cytoplasmic compartments were calculated for background corrected N/C ratios per cell and log transformed. Finally, average N/C ratios were determined per cell line.

2.2.6 Specific detection of DNA

2.2.6.1 Surveyor[®] mutation screening

To assess the genotypes of CRISPR/Cas9 genome edited cells, Surveyor[®] mutation screenings were performed. This involved PCR amplification of CRISPR/Cas9-targeted sequences using 2X PCR Master Mix, Surveyor[®] assay-designated oligo pairs (Table 2.2), and 50ng genomic DNA as template. Additionally, genomic template between parental cells (293T or Sf9 as indicated) was mixed with selected CRISPR/Cas9 genome edited cell line at the indicated ratios of 50ng total genomic DNA template, and PCR amplified. PCR of target sequences from parental genomic DNA served as reference control for subsequent mutation screenings. Therefore, Surv_sgRNA1.F(oward) and R(everse), Surv_sgRNA2.F(oward) and R(everse), Surv_sgRNA3.F(oward) and R(everse), and Surv_sgRNA4.F(oward) and R(everse) oligo pairs targeted the *YBX1*sgRNA1-4-specific sequences, respectively, in selected mammalian cell lines. Alternatively, the *SFYB*sgRNAsf1-targeted sequence in Sf9-based cell line was amplified using Surv-sgRNAsf1.F(oward) and R(everse) oligos. After confirmation of approximate sized PCR amplicon(s) by 2% TBE agarose gel electrophoresis and visualisation and imaging using a UV trans-illuminator, 20µL of sample was subjected to hybridisation, to facilitate heteroduplex formation between wildtype sequences and CRISPR/Cas9-mutated sequences. Heteroduplex formation was performed on a Biorad PTC-200 DNA Engine Thermal Cycler, and the programmed thermocycling conditions set up: 95°C for 10mins; -2.0°C/sec ramp down from

95°C-85°C, 85°C for 1min; -0.3°C/sec ramp down from 85°C-75°C, 75°C for 1min; -0.3°C/sec ramp down from 75°C-65°C, 65°C for 1min; -0.3°C/sec ramp down from 65°C-55°C, 55°C for 1min; -0.3°C/sec ramp down from 55°C-45°C, 45°C for 1min; -0.3°C/sec ramp down from 45°C-35°C, 35°C for 1min; -0.3°C/sec ramp down from 35°C-25°C, 25°C for 1min; and hold at 4°C. Mutation screening was then performed following the Surveyor® Mutation Detection Kit (IDT) instructions by taking 10µL of hybridised DNA and adding 0.1µL of 0.15M MgCl₂, 2µL Surveyor® Enhancer S (composition – propriety information), and 1µL of Surveyor® Nuclease S. Reactions were gently spun down briefly, and then incubated at 42°C for 1h to catalyse nuclease S-mediated digestion of mismatched DNA, which would be present only in heteroduplex DNA. Finally, 3µL of Surveyor® Stop Solution (composition – propriety information) was then added to terminate the digestion reaction. Samples were immediately mixed with 6X DNA Loading Dye and subjected to 2% TBE agarose gel electrophoresis, to identify digestion bands and determine genotypes.

2.2.6.2 High resolution melting (HRM) curve analyses

To further confirm the genotype of CRISPR/Cas9-edited cells CRISPR/Cas9-targeted sequences were subjected to HRM, as a sensitive measure to discriminate against wildtype sequences in a cloned cell population. Therefore, PCR amplified using 2X PCR Mastermix, Surveyor® assay-designated oligo pairs as described in 2.2.6.1. PCR amplicons were gel purified using the MinElute Gel Extraction Kit (Qiagen) after 1% TAE agarose gel electrophoresis, and DNA concentration quantified by Nanodrop™ 1000. Parental 293T genomic DNA served as reference control for subsequent HRM curve reactions and analysis. HRM reactions were prepared using SensiFAST™ HRM Kit (Bioline) and comprised of the following at their final concentrations: 1X SensiFAST™ HRM Mix constituting DNA polymerase, dNTPs (dATP, dGTP, dCTP, and dTTP), EvaGreen® dye, and 3mM MgCl₂, HRM designated forward and reverse oligo pairs at 0.4µM each, and 2µL template DNA, made to 20µL with nuclease-free H₂O. Non-template control reactions were prepared in parallel, with the substitution of template DNA with nuclease-free H₂O. HRM-PCR reactions were performed using a Rotor-Gene 6000 platform (Qiagen) and Rotor-Gene Q Series Software (Qiagen). Thermocycling conditions were set up as follows: 95°C for 3mins; followed by 45 cycles of 95°C for 5secs, 58°C for 10secs, and 72°C for 10secs with acquisition. HRM followed at 0.1°C/step ramp up from 70°C-95°C with 90secs wait of pre-melt conditioning on the first step, and 1sec wait for all other steps, with acquisition. Dissociation curves derived from parental 293T template were assigned as homozygous control for genotype analyses. Dissociation peaks

were annotated and their associated dissociation temperatures set with the limit of $\pm 0.5^{\circ}\text{C}$. HRM curve analyses were set to the 90% confidence threshold. HRM-PCR reactions were also subjected to 2% TBE agarose electrophoresis to confirm PCR amplification or target DNA.

2.2.6.3 Quantitative (q)PCR

5 μL of crude rAAV vector or purified rAAV vector was first treated with 50U of Benzonase[®] Nuclease for 1h at 37 $^{\circ}\text{C}$. This was followed by treatment of samples with 50U of DNase I (Invitrogen) in DNase I Buffer made to a final 1X (5X: 100mM Tris-HCl [pH 6.8], 10mM MgCl₂, 250mM KCl) in 50 μL reaction volume. Samples were treated with DNase I overnight at 37 $^{\circ}\text{C}$, following which samples were briefly spun down and nucleases were heat-inactivated at 75 $^{\circ}\text{C}$ for 20mins. Samples were briefly spun down and then placed on ice. 1 μL of 20mg/mL Proteinase K (Roche) was then added, and reactions incubated at 50 $^{\circ}\text{C}$ for 2h for Proteinase K to degrade rAAV capsid proteins and liberate rAAV vector genomes. Proteinase K was then heat-inactivated at 95 $^{\circ}\text{C}$ for 20mins, and samples briefly spun down. 5 μL of sample was then used to measure physical rAAV2 vector genome titres by qPCR in duplicate, using oligos and Taqman probes specific for the CMV promoter – CMV.F(oward), CMV.R(everse), and CMV.hyd_probe (Table 2.2). Non-template control reactions for qPCR assays were also included and involved the addition of H₂O instead of rAAV2 sample, in duplicate. Plasmid pAAV2-hrGFP was used to generate a standard curve by ten-fold serial dilution that ranged from 10⁸-10¹ copies (5 μL each in duplicate) using nuclease-free H₂O. PCR reactions were performed using LightCycler[®] 480 Probes Master (Roche), and composed of the following at their final concentrations: 1X reaction buffer with FastStart Taq DNA polymerase, dNTP mix, 3.2mM MgCl₂, and 5 μL of testing sample, made to 20 μL using nuclease-free H₂O. Reactions were assembled in white LightCycler[®] 480 Multiwell Plate 96. The LightCycler[®] 480 instrument II platform (Roche) was used for qPCR and thermocycling conditions were set as follows: 10mins, 95 $^{\circ}\text{C}$ initial denaturation; followed by 45 cycles: 15secs denaturation at 95 $^{\circ}\text{C}$, 30secs annealing at 60 $^{\circ}\text{C}$, and 5secs extension at 72 $^{\circ}\text{C}$, with acquisition. Absolute quantitation of vector genome titres was performed using the LightCycler[®] 480 instrument II with automated thresholding and its pre-programmed fit-points algorithm. Recombinant AAV titres were calculated as vector genomes/mL from the standard curves.

2.2.7 Pulldown and protein binding assays

2.2.7.1 Double-stranded DNA-affinity YB1 pulldown

5'-desthiobiotin modified capture probes were PCR amplified using Promega 2X Master Mix (see Table 2.6 for list of oligos for PCR). Using pAAV2-MCS as template the *ITR* capture probe was amplified with *ITR_YB1.F(orward)* and *R(everse)* oligos. Control capture probes also used pAAV2-MCS and L- or R- *ITR_YB1.F(orward)* and *R(everse)* oligo pairs, for left- and right- *ITR* flanking sequences, respectively. Using pHelper as template the *rep* capture probe was amplified with *rep_YB1.F(orward)* and *R(everse)*. Using pAAV2/2-RC as template the *E2A* capture probe was amplified with *E2A_YB1.F(orward)* and *E2A_YB1.R(everse)*. All PCR amplicons were purified by 1% TAE agarose gel electrophoresis and gel extraction. Otherwise, short dsDNA 3'-biotin labelled capture probes were generated by annealing *ITR(-)sense.F(orward)* and *R(everse)*, *E2A(-)sense.F(orward)* and *R(everse)*, *rep(+)**sense.F(orward)* and *R(everse)*, *GCcomp.F(orward)* and *R(everse)*, and *Luciferase.F(orward)* and *R(everse)* oligo pairs, together.

Dynabeads® M-270 Streptavidin Beads (Invitrogen) were initially vortexed for 30secs for complete resuspension of beads. Streptavidin beads were transferred to clean Eppendorf tubes, and thrice washed in bulk with 500µL of 0.5X SSC (20X: 3M NaCl, 0.3M sodium citrate) and magnetic rack for 2mins each. Streptavidin beads were then washed thrice with 500µL Binding Buffer A (25mM HEPES [pH 7.5], 100mM KCl, 12.5mM MgCl₂, 1mM DTT, 0.1% IGEPAL-CA630 [v/v], 20% glycerol [v/v], 3% BSA) and 2mins on magnetic rack, followed by incubation in 500µL of Binding Buffer A for 30mins at RT. The Binding Buffer A was aspirated after exposure to a magnetic rack for 2mins, and beads were resuspended in their original volume using Binding Buffer A. Where indicated, 1mg (100µL) or 0.2mg (20µL) of streptavidin beads was used per pulldown.

10µg of 5'-desthiobiotin-labelled DNA capture probes was then added to 1mg beads and mixed to a final volume of 500µL of Binding Buffer A. As a negative control for protein pulldown, streptavidin beads were incubated with Binding Buffer A only. Capture probes were immobilised for 30mins at RT with rotation. Coated beads were then isolated by magnetic rack for 2mins and free capture probe aspirated from the supernatant. Beads were washed thrice with 500µL of Binding Buffer A and magnetic rack for 2mins. After final aspiration of supernatant, 200-500µg (as indicated) of cell lysate (lysed in Lysis Buffer A) and 5µg poly(dI-dC) (Sigma-Aldrich) was mixed and volumes normalised to 500µL in ice-cold Lysis Buffer A.

Protein binding was facilitated overnight at 4°C with rotation. Alternatively, 2µg (approximately 100pmol) of 3'-biotin labelled dsDNA capture probe (corresponding to 100µL of annealed oligo pairs; see section 2.2.1.7) was added to 0.2mg of beads, and final volume adjusted to 500µL with Binding Buffer A. 3'-biotin labelled capture probes were immobilised for 30mins at RT with rotation, followed by three washes using Binding Buffer A and magnetic rack for 2mins. This was followed by the addition of 100µL of cold competitor DNA probe (see section 2.2.1.7) to DNA capture probe-bound beads at the required ratio. Negative controls included the omission of the cold competitor dsDNA probe, and volume was substituted by Binding Buffer A. 200µg of cell lysate (lysed in Lysis Buffer A) was also added per pulldown, and final volumes adjusted to 500µL of Binding Buffer A. Reactions were incubated at 4°C for 1h with rotation. In both cases the streptavidin beads were washed five-times with ice-cold 500µL of Binding Buffer A and twice with ice-cold 1X PBS, via magnetic rack for 2mins each. Beads were then mixed with 50µL 5X Laemlli Buffer and boiled for 10mins at 95°C to dissociate capture probe-protein from streptavidin beads. A final pulldown of beads was performed using a magnetic rack for 2mins, and supernatant transferred to a new Eppendorf tube. Samples were boiled again for 5mins at 95°C before loading 20µL of sample onto a 10% SDS-PAGE gel and subjected to Western blotting for YB1 pulldown.

2.2.7.2 Single-stranded DNA-affinity YB1 pulldown

50pmol of 3'-biotin labelled ssDNA capture probe (Table 2.6) was diluted to 500µL of Binding Buffer A and added to 0.2mg of streptavidin beads. Capture probes were immobilised for 30mins at RT with rotation. This was followed by thrice washing the beads using Binding Buffer A and magnetic rack for 2mins. The 100µL of cold competitor ssDNA probe – GCcomp.F(oward) was then added, and corresponded to the required ratio of cold competitor by scaling-up the volume of 10µM cold competitor ssDNA probe. Controls included no addition of cold competitor ssDNA probe (1:0 ratio of ssDNA capture probe:cold ssDNA competitor); the volume of which was substituted by Binding Buffer A. 200µg of cell lysate (lysed in Lysis Buffer A) was added per pulldown, and final volumes adjusted to 500µL of Binding Buffer A. Reactions were incubated at 4°C for 1h with rotation. Beads were then washed with 500µL Binding Buffer A five times and 500µL 1X PBS twice, using a magnetic rack for 2mins. Upon the final wash beads were pelleted down at 13,000rpm for 5mins and then exposed to magnetic rack for 2mins for the complete aspiration of the final wash volume. Beads were resuspended in 50µL of 5X Laemlli Buffer and boiled for 10mins at 95°C to dissociate capture probe-protein from streptavidin beads. Beads were finally pelleted down

for 1min at 13,000rpm and then supernatant harvested after using a magnetic rack. The supernatant was transferred to a new Eppendorf tube, then boiled again for 5mins at 95°C before loading 20µL of sample onto a 10% SDS-PAGE gel and subjected to Western blotting for YB1 pulldown.

2.2.7.3 Electrophoretic mobility shift assays (EMSA)

Selected plasmids (200ng) or dsDNA capture probes (50ng) were incubated with purified rSfYB(His)₁₀ protein at various quantities (as indicated), and were prepared by mixing each component in Cutsmart Buffer (NEB) made to a final 1X concentration in volume of 30µL in nuclease-free H₂O. As a negative control, each plasmid or dsDNA capture probe was prepared in Cutsmart Buffer without the addition of rSfYB(His)₁₀ protein. The volume of which was substituted with 1X Cutsmart buffer. Binding reactions were subjected to 37°C incubation for 30mins and then cooled on ice immediately. Samples were then mixed with 6X DNA Loading Dye and then subjected to agarose gel electrophoresis, and gels visualised and imaged.

2.2.8 Statistics

Graphs and statistical analyses were performed with GraphPad Prism 6 Software. Student's *T*-test was used to assess for statistical significance. Growth curves and LDH cytotoxicity assays were analysed using two-way ANOVA with Sidak test for multiple comparisons. Data is presented as mean ± standard deviation (±SD), with *P* < 0.05 as the upper cut-off for statistical significance. *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001 or *P* < 0.0001 (as indicated).

Chapter 3: Establishment of YB1 Knockout Cell Lines for Recombinant AAV Vector Production Using CRISPR/Cas9 Genome Editing

3.1.0 Introduction

Recombinant Adeno-associated viral (rAAV) vectors have become a highly attractive and successful vector tool for gene therapy applications. This is due to a number of inherent properties associated with AAV that make this the vector of choice for targeting genetic diseases. Advantages include the fact that AAV can mediate long-term transgene expression *in vivo* (Xiao *et al.*, 1996), and rAAV vectors have shown clinical efficacy in a number of clinical trials (Nathwani *et al.*, 2011; Nathwani *et al.*, 2014; French *et al.*, 2018; Miesbach *et al.*, 2018). More promising is the fact that the first EU-licensed gene therapy product, Glybera, and the even more recently US-approved Luxturna, are both rAAV-based gene therapy products for the treatment of hereditary lipoprotein lipase deficiency (Carpentier *et al.*, 2012) and Leber's congenital amaurosis (Bennett *et al.*, 2016), respectively. These collectively demonstrate that rAAV-based gene therapies are a promising advancement as a modern healthcare and medicine.

However, the clinical efficacy of using rAAV vectors is restricted by certain limitations. The prevalence for neutralising antibodies against different AAV serotypes in the human population, for example, presents as a significant obstacle towards a completely efficacious gene therapy (Kotterman *et al.*, 2015). Additionally, the demand for sufficient rAAV vector quantities is constrained by current production systems. There still remains a need for rAAV vector quantities that meet the demands of clinical studies; a situation worsened by the strategy to circumvent the limitation of pre-existing immunity by administration of high titres of rAAV as one of the ways to improve the effectivity of rAAV vector products. Particular advancements in rAAV vector designs have been explored, which encourage further engineering rAAV vectors. Approaches in this regard include rAAV vectors that elicit reduced, or ablate, immune responses (Martino *et al.*, 2013; Wang *et al.*, 2014; Tse *et al.*, 2017), exhibit improved transduction profiles or have enhanced target specificities (Aslanidi *et al.*, 2013;

Buning *et al.*, 2015; Ling *et al.*, 2016). AAV vector production methods have also been further developed from the conventional triple plasmid co-transfection of 293T cells (Xiao *et al.*, 1998), or co-infection of Sf9 cells with recombinant baculoviruses (Urabe *et al.*, 2002; Smith *et al.*, 2009). These developments include the generation of partially stable cell lines for *rep* and/or *cap* expression, therefore off-setting the limitation of exogenously introducing these proteins into the production cycle (Gao *et al.*, 2002b; Mietzsch *et al.*, 2014). However, our understanding of the role of host cellular factors towards AAV genome processing and/or AAV assembly remains limited. Recently, host cell factors, including human Y-Box protein (YB1) protein (encoded by *YBX1*), have been found associated with rAAV particles; and targeted shRNA-mediated knockdown of YB1 protein expression in 293T cells correlated with enhanced rAAV vector titres – up to 45-fold more rAAV2 vector genomes were quantified compared to scramble controls (Satkunanathan *et al.*, 2014). This approach demonstrated a rather novel rAAV production system involving the regulation of endogenous, cell-intrinsic protein expression to influence rAAV vector production.

In the present study, host cell gene expression was modified permanently using the RNA-guided clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing technology. Particularly, host cell genome editing focused on targeting the human *YBX1* gene, given the targeted knockdown of YB1 protein expression correlated with an enhanced rAAV vector producer cell line after transient triple transfection for rAAV2 and 8 production (Satkunanathan *et al.*, 2014). Additionally, the enhanced rAAV vector producer cell line quality was lost over time, presumably due to the progressive loss of YB1 downregulation by shRNA (Satkunanathan *et al.*, 2014). Ultimately, CRISPR/Cas9 offers the ability to permanently edit the host genome, with minimal changes to its entire architecture and composition of the target genome. This is in contrast to alternative methods of gene manipulation, such as shRNA technology which requires the incorporation of shRNA and viral vector elements randomly into the genome in order to maintain their effect. Additionally, shRNA or RNAi mediated gene silencing strategies are often complicated by off-target effects including the upregulation of rhodopsin or interferon promoter activities (Masuda *et al.*, 2016), or transient functionality, with eventual silencing or loss of shRNA expression over time (Ahn *et al.*, 2011; Satkunanathan *et al.*, 2014). Therefore, use of CRISPR/Cas9 technology was harnessed to offset the transient shRNA-mediated downregulation of YB1 expression experienced by Satkunanathan *et al.* (2014), to potentially develop a permanent high titre rAAV vector production platform.

The aim of this study was to utilise the CRISPR/Cas9 genome editing system to target the *YBX1* gene and knockout YB1 protein expression in 293T cells. Further aims included an extensive characterisation of the YB1 knockout cell lines generated by CRISPR/Cas9 genome editing. Additionally, the suitability of the approach to establish stable YB1 knockout cell lines for the intended purpose of rAAV vector production was assessed. The data demonstrated that targeted knockout of YB1 protein expression was successfully achieved in 293T cells, and supported by a range of both phenotypic and genotypic assessments. Although, YB1 knockout did not enhance rAAV2 vector titres, specific targeting of YB1 in 293T cells generated an undesirable phenotype with regards to transient transfection modalities. In particular, the use of chloroquine and Calcium Phosphate precipitation methodology had an adverse effect of YB1 knockout, which potentially limits the broader use of these YB1 knockout cell lines.

3.2.0 Results

3.2.1 Cloning *YBX1*-specific gRNA designs into GeneArt CRISPR nuclease vector

A total of four gRNAs were designed that targeted the *YBX1* gene at different sequences (see Fig. 3.1A for a schematic of gRNA target sequences relative to the entire *YBX1* gene structure). These were annotated *YBX1*sgRNA -1, -2, -3, and -4, and were specific targeting exon 1, 5, intron 6, and exon 7 of the *YBX1* gene, respectively. The gRNA designs were each cloned into a pre-linearised GeneArt™ CRISPR nuclease vector from Invitrogen as their complementary dsDNA intermediate as per manufacture's guidelines (Fig. 3.1B). Sequencing data showed that each *YBX1*sgRNA(1-4) DNA sequence had been successfully cloned into the all-in-one vector as intended (Figs. 3.1C-F; underlined sequences refer to the intended gRNA target sequences). Further to this, sequence reads upstream and downstream of the cloned gRNA sequences corresponded to the GeneArt™ CRISPR nuclease vector backbone. Therefore, CRISPR nuclease vectors with the required *YBX1*-target specific sgRNA were successfully constructed with the intention of targeting different intron or exon targets of the *YBX1* locus in 293T cells. CRISPR vectors plasmids were referred to as pCRISPR-*YBX1*sgRNA1-4 for their respective *YBX1*sgRNA design (gRNA1-4).

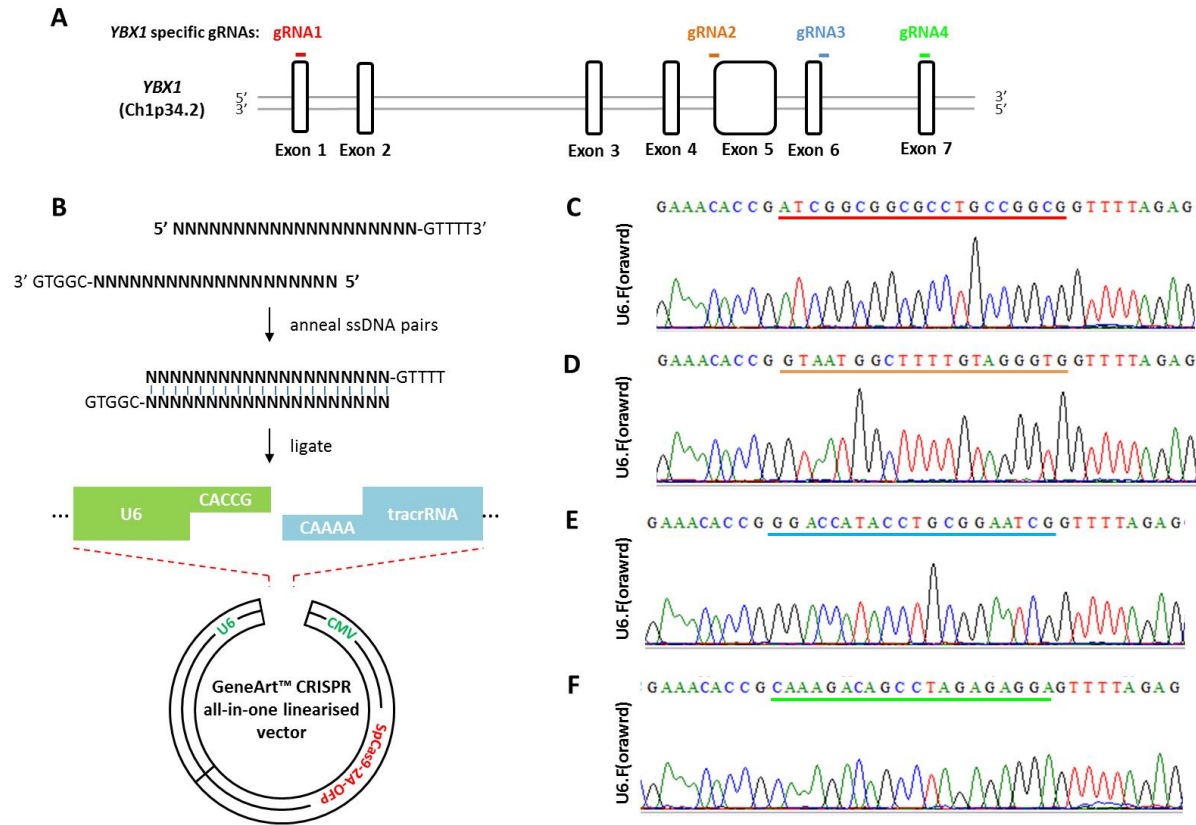


Figure 3.1 Cloning of YBX1-specific into GeneArt™ linearised, all-in-one, CRISPR nuclease vector for CRISPR/Cas9 genome editing. **A**) Schematic showing YBX1-specific gRNA designs and their target sequence location relative to the entire YBX1 gene structure. **B**) Cloning strategy to clone YBX1sgRNA(1-4) sequences into the GeneArt™ linearised, all-in-one CRISPR nuclease vector involving annealing ssDNA pairs with appropriate overhangs. Annealed oligos encode the YBX1-specific gRNA(1-4). These are individually ligated as a dsDNA intermediate into the GeneArt linearised vector downstream of a U6 promoter and upstream to the tracrRNA encoded sequence. **C-D**) Sequencing and sequencing chromatograms confirmed cloning of GeneArt™ CRISPR nuclease vector harbouring **(C)** YBX1sgRNA1 sequence, or **(D)** YBX1sgRNA2 sequence, or **(E)** YBX1sgRNA3 sequence, or finally **(F)** YBX1sgRNA4 sequence. Underlined sequences correspond to the intended gRNA design.

3.2.2 Initial screening of transfected 293T cells identified potential YB1 knockout clones

To establish YB1 knockout cell lines, 293T cells were transfected with pCRISPR-*YBX1*sgRNA1-4. The transfection efficiency of pCRISPR-*YBX1*sgRNA1-4 plasmids was examined by fluorescent microscopy given the co-expression of orange fluorescent protein (OFP) reporter with SpCas9. Overall transfection efficiency was empirically determined, and almost all cells were positive for OFP expression for each vector construct used (Fig. 3.2A; panels a-b, c-d, e-f, and g-h, for pCRISPR-*YBX1*sgRNA1-4 transfections, respectively). Non-transfected 293T cells served as a negative control for background OFP expression (Fig. 3.2A, panels i-j), and exhibited no OFP expression by fluorescence microscopy. However, transfection efficiency and OFP expression does not reflect genome editing. In order to establish a cell line that is truly a knockout of YB1, transfected 293T cells underwent single cell cloning and five clones of each target was subjected to initial screening by Western blot against bulk- and non- transfected 293T control.

The bulk transfected cells for each *YBX1*-specific sgRNA indicated that gene editing had been achieved using CRISPR/Cas9 genome editing. YB1 signal was reduced for bulk transfected cell populations when compared to 293T control cells (Figs. 3.2B-E, see lanes 2 and 3). However, this was not a consistent observation for *YBX1*sgRNA2 between 293T bulk transfected cells in duplicate (Fig. 3.2C, compare lanes 2 and 3) relative to control 293T. Bulk set 1, instead, demonstrated equivalent YB1 protein levels compared to non-transfected 293T control, indicating little to no gene editing was evident in this particular bulk transfected population. Analysis of single cell clones (five for each pCRISPR-*YBX1* sgRNA1-4 transfection) indicated that targeting exon 1 and intron 6 produced the most YB1 knockout clones, contributing to 4/20 (20%) clones analysed (Figs. 3.2B and D, and lanes 2 and 5-7, respectively). Overall, targeting exon 1 using *YBX1*sgRNA1 demonstrated 20% (1/5 clones) knockout rate (Fig. 3.2B), targeting intron 6 using *YBX1*sgRNA3 produced clones that showed 60% (3/5 clones) knockout rate (Fig. 3.2D). Targeting exon 5 using *YBX1*sgRNA2 also yielded mostly marked reduction of YB1 levels in single cell clones analysed (Fig. 3.2C, lanes 5-7). Similarly, targeting of exon 7 using *YBX1*sgRNA4 produced one clone with a reduction in YB1 levels post-single cell cloning (Fig. 3.2F, lane 4). Unusually, a single clone exhibited YB1 overexpression when compared to the equal loading of 293T control (Fig. 3.2F, lane 7). No single cell clones exhibited an YB1 knockout phenotype for clones derived from targeting exon 5 or 7. Clones A2, B2, C5, and D1 (the prefix A-D refers to *YBX1*-sgRNA1-4 targeting, respectively) were selected for further characterisation.

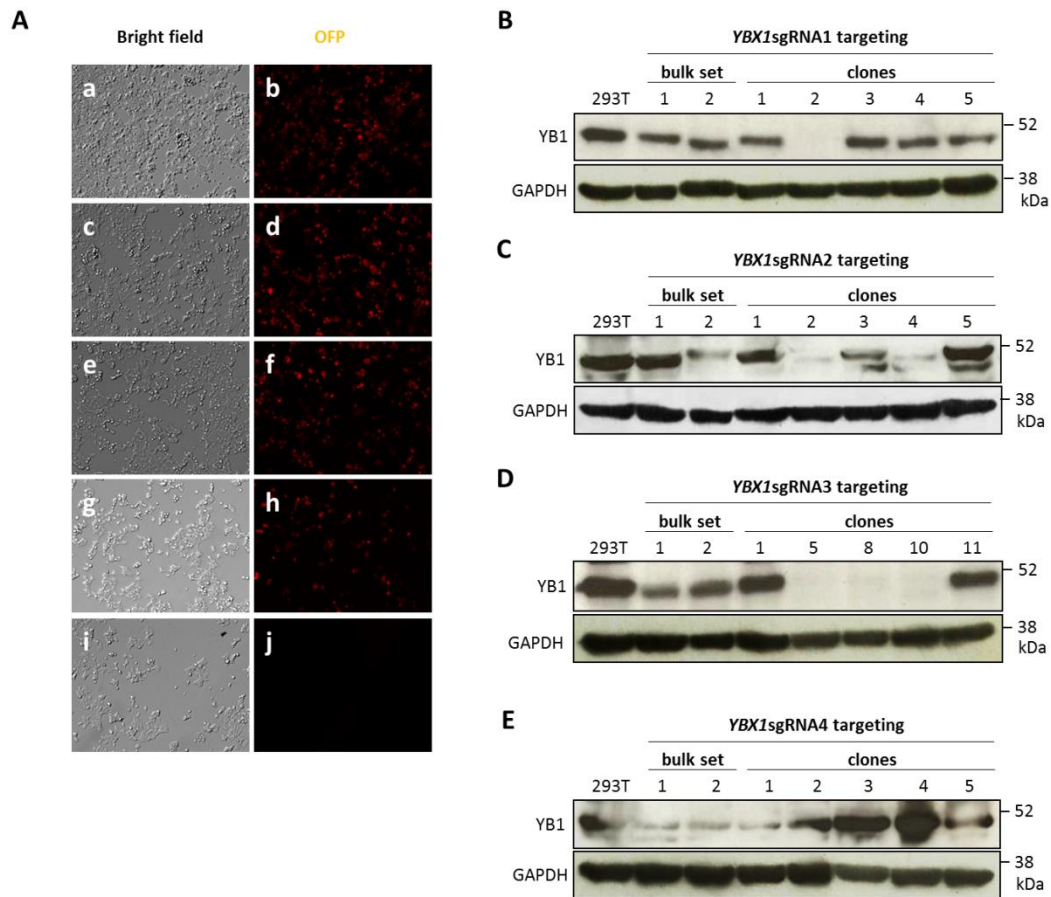


Figure 3.2 Establishing YB1 knockout cell lines. **A**) Bright field and fluorescent microscopy images (left and right panels, respectively) indicated that most cells exhibited GFP expression for pCRISPR-YBX1sgRNA1-4-transfected 293T cells, 72h post-transfection (**panels a-b, c-d, e-f, and g-h, respectively**). Non-transfected 293T cells served as negative control for transfection (**panels i-j**). **B-E**) Western blot using α YB1 (ab76149) of non-transfected 293T (baseline control for YB1 expression; lane 1), bulk transfected 293T populations (lanes 2 and 3), and five single cell clones (lanes 4-8) were analysed for each pCRISPR-YBX1sgRNA1-4 transfections, respectively. Western blot analysis indicated that partial knockout and complete knockout cell lines were achieved. GAPDH protein expression was used as loading control (bottom panels for each).

3.2.3 Screening YB1 knockout clones for epitope disruption as a result of CRISPR/Cas9 genome editing of YBX1

Depending on the gene editing by CRISPR/Cas9, mutations that disrupt the target epitope recognised by α YB1 (ab76149) may have interfered with Western blot analysis. Therefore, further Western blotting using α YB1 raised against different N- (PA5-19453) and C-terminus (ab114999) epitopes was assessed. Overall, Western blot analysis showed comparative profiles were observed between the three different antibodies for clones A2, B2 and C5 (Fig. 3.3A). Although difficult to observe in the blots (Figs. 3.3A and C) a diminished YB1 signal is observed for B2 clone for each α YB1 (see black arrows). However, using alternative α YB1 (ab114999 and PA5-19543), clone D1's YB1 levels were considered comparable to the 293T control. These findings suggest that genome editing in D1 clone may have disrupted the

epitope for ab76149 only. Furthermore, the *YBX1*sgRNA4 design targeted *YBX1*'s last exon, of which subsequent genome editing may have occurred so far into the *YBX1* gene sequence that the gene may have mostly retained in-frame coding sequences and significant homology to wildtype *YBX1*. Sufficient homology may enable the avoidance of non-sense mediated mRNA decay, with the ab76149-specific epitope disrupted only. The N-terminus-specific α YB1 (PA5-19453) was raised against peptide corresponding to 1-100aa of YB1, and C-terminus-specific α YB1 (ab114999) was raised against peptide corresponding to 224-274aa. With the C-terminus specific α YB1 (ab76149) epitope target undisclosed by its manufacturer, it becomes difficult to further analyse the exact nature of this potential 'false-positive' partial knockout. Given the ambiguous YB1-expression profile exhibited by clone D1 this clone was excluded from further characterisation.

Clones A2, B2, and C5 were then subjected to a second round of single cell cloning. The purpose of this was to analyse subclones to reaffirm that the selected clones were not heterogeneous for the YB1-expressing phenotype, and that the initial round of single cell cloning had been efficient. The second generation of A2 and C5 clones (subclones A2.1-2.3 and C5.1-3, respectively) demonstrated similar profiles to their parental YB1 knockout cell lines (Fig. 3.3B). This confirmed confidence in single cell cloning technique as producing *bona fide* knockouts. Clone B2 demonstrated a consistent reduction in YB1 when compared to the 293T control, as did its subclones (B2.1-3) when using the different α YB1 (Fig. 3.3C). Western blotting A2, B2 and C5 subclones with a larger coverage of YB1-specific antibodies provided evidence that the lack of YB1 signal observed in A2 and C5 clones, and their subclones, was likely due to the complete knockout of the *YBX1* gene by CRISPR/Cas9 genome editing rather than epitope disruption.

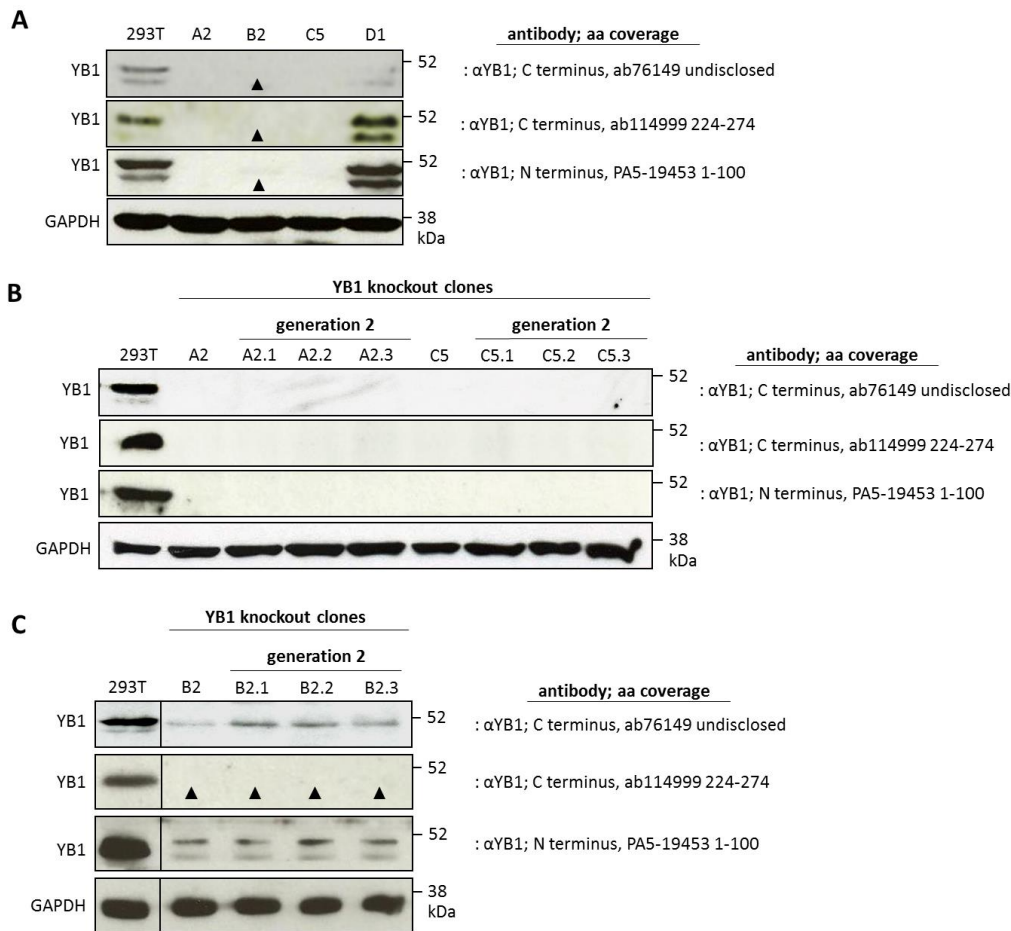


Figure 3.3 Western blot analysis for epitope disruption of selected YB1 knockout cell lines. Epitope disruption was ruled out for the observed YB1 knockout phenotypes, especially for clones A2, B2 and C5. **A)** Western blot analysis of selected clones: A2, B2, C5 and D1 using three YB1-specific antibodies (ab76149, ab114999, and PA5-19453) that bind to different YB1 epitopes. D1 clone demonstrated epitope disruption, which resulted of the partial knockout phenotype previously observed. Western blot analysis of selected clones **(B)** A2, C5, and **(C)** B2, and their corresponding sub-clones (generation 2), showing that the knockout and partial knockout phenotypes observed were considered preserved and monoclonal. Non-transfected 293T was used as a base-line control. GAPDH expression was used as a loading control. Black arrows refer to diminished YB1 signal observed for B2 clone, which were difficult to observe.

3.2.4 Phenotyping YB1 knockout cell lines using immunocytochemistry (ICC)

The YB1 knockout phenotype was further confirmed using immunocytochemistry. The benefit here was exploring for YB1 knockout phenotypes in selected clones, using α YB1 to recognise whole, intact YB1 molecule, as opposed to SDS-reduced protein lysates as per Western blotting. Essentially, the exposure of epitopes is fundamentally different – whole molecule in its native state versus linearized polypeptide (Fieser *et al.*, 1987). 293T cells provided the positive control for YB1 expression and localisation by ICC; the localisation of which was defined as cytoplasmic as no YB1 signal was identified in the nuclear compartment, which was differentiated by DAPI-staining (Fig. 3.4A, panel a-c). YB1 was undetectable for selected YB1 knockout clones A2, B2, and C5, reflecting the YB1 knockout phenotypes observed by Western

blotting (Fig. 3.4A, panels d-f, g-i, and j-l, respectively). Clone B2 previously showed a partial knockout phenotype by Western blotting, but using ICC, clone B2 displayed profiles similar to that of A2 and C5 clones (Fig. 3.4A, panels g-i). Control 293T cells stained with either α YB1 (PA5-19453) or secondary antibodies only functioned as negative staining controls, and demonstrated only background fluorescence (Fig. 3.4A, panels m-o and q-s, respectively).

Orthogonal cross-sections of 293T and selected YB1 knockout cells confirmed the cytoplasmic localisation and knockout of YB1 in 293T cells and selected cell clones, respectively (Figs. 3.4B-E). Taking into account the depth of the cell (z-plane) the localisation of endogenous YB1 was further confirmed as cytoplasmic in parental 293T cells (Fig. 3.4B), and undetectable throughout YB1 knockout clones A2, B2, and C5 (Figs. 3.4C-E). The degree of fluorescent intensity, reflecting YB1 abundance, was quantitatively analysed. Fig. 3.4F represents the normalised or corrected total cell fluorescence (CTCF) of 293T, A2, B2 and C5 cell lines against negative staining of control (293T secondary Abs). CTCF measurements indicated that A2, B2, and C5 clones, especially B2 clone, exhibited similar average CTCF (525.6 ± 148.9 , 1078.53 ± 351.41 , and 698.1 ± 136 , respectively) to 293T negative staining control (293T secondary Abs; 955.86 ± 230.22). The reduced average CTCF measurements were considered statistically significant for A2 and C5 ($n = 20$ individual cells measured; $P < 0.001$, $P < 0.01$). Nonetheless, CTCF in 293T positive control for YB1 staining was markedly measurable (7969.49 ± 2621.94), and statistically significant to 293T negative staining control ($n = 20$ cells each, $P < 0.001$), and by extension the selected YB1 knockout clones (Fig. 3.4F).

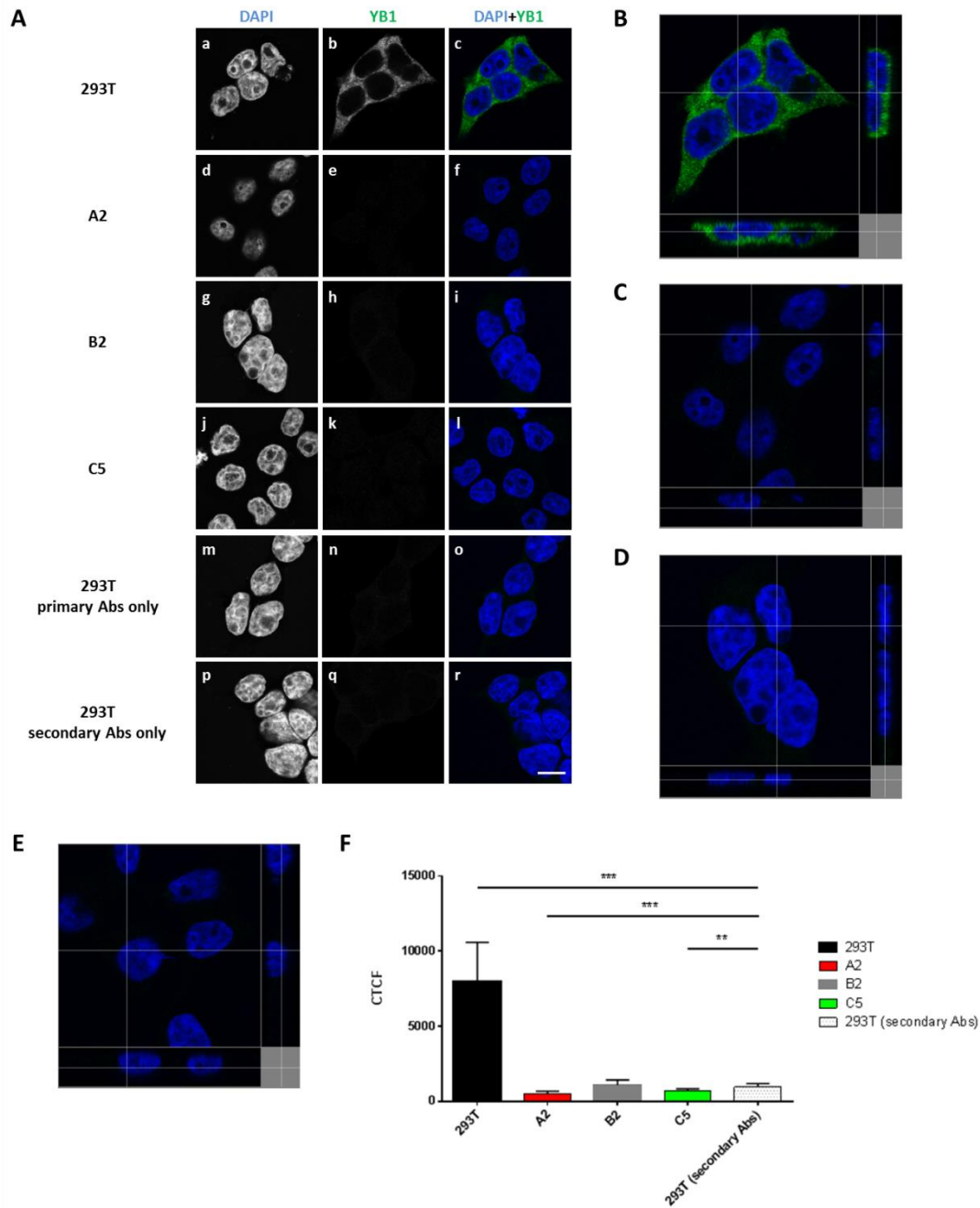


Figure 3.4 ICC and confocal microscopy analysis for YB1 knockout phenotype. YB1 knockout phenotypes were further validated using ICC and quantifying CTCF and compared between 293T, A2, B2, C5. **A)** ICC analysis by confocal laser scanning microscopy demonstrated that YB1 was only detectable for 293T positive control (**panels a-c**). Panels (**d-f**), (**g-i**), and (**j-l**) refer to selected YB1 knockout clones (A2, B2, and C5, respectively) stained with both primary (PA5-19453) and secondary (α Rabbit-Alexa Fluor 488) antibodies for YB1 detection. Panels (**m-o**) and (**p-r**) refer to 293T cells stained with either primary or secondary antibodies only, respectively, and functioned as staining controls for background fluorescence. Scale bar = 20 μ m, and representative of all panels. **(B)** Orthogonal section analysis of 293T stained cells for YB1 expression and localisation, demonstrating that YB1 protein is almost exclusively localised to the cytoplasm and absent from the nucleus. Whereas, orthogonal section analysis of YB1 knockout clones **(C)** A2, **(D)** B2, and **(E)** C5 demonstrated an absence of YB1 signal throughout the cell, considering all planes of view. Figure legend continues next page.

Figure 3.4 ICC and confocal microscopy analysis for YB1 knockout phenotype. F) Graph presenting CTCF measurements between 293T positive control, A2, B2, and C5, and 293T secondary Abs only stained negative control, indicating that YB1 knockout clones exhibited fluorescence signal similar to the 293T secondary Abs staining control. Error bars reflect \pm SD from mean, $n = 20$ measured cells for each cell line, except 293T secondary Abs only stained control, $n = 5$; **, $P < 0.01$; ***, $P < 0.001$ or $P < 0.0001$. Abs, antibody; CTCF, corrected total cell fluorescence.

3.2.5 Sequencing YB1 knockout clones demonstrated mutations by CRISPR/Cas9

The selected YB1 knockout cell clones were also analysed by genomic methods to establish their exact genotypes. Genomic DNA was isolated from parental 293T cells, following which the equivalent regions targeted by *YBX1*sgRNA1-4 were PCR amplified as 400-500bp amplicons using Surv_gRNA1, 2, 3, and 4, forward and reverse pairs (Figs. 3.5A and B; referred to as 293A-D). The intended genomic regions from control 293T, reflecting regions targeted centrally by *YBX1*sgRNA1 (expected 407bp, lane 293A), *YBX1*sgRNA2 (expected 467bp, lane 293B) amplicons were successfully identified (Fig. 3.5A). Likewise, genomic regions reflecting the regions targeted centrally by *YBX1*sgRNA3 (expected 405bp, lane 293C) and *YBX1*sgRNA4 (expected 435bp, lane 293D) were also successfully amplified. The required sequences could be PCR amplified in 293T and selected YB1 knockout cell lines for subsequent purification, despite non-template control reactions (NTC-A-D) for each PCR amplification showing formation of primer dimers only (Figs. 3.5A and B).

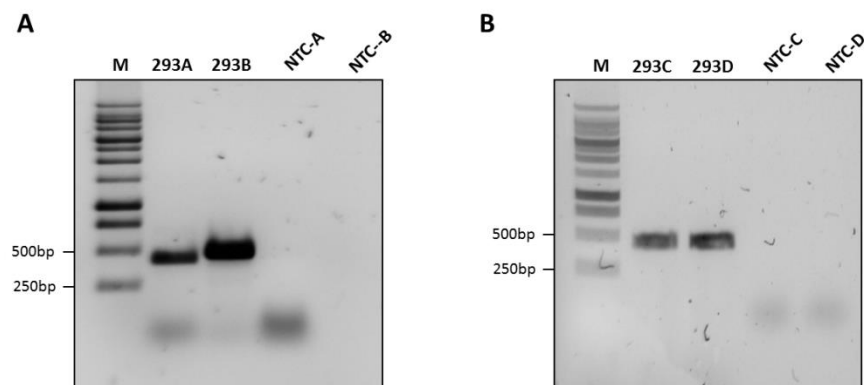


Figure 3.5 PCR amplification of target sequences from 293T genomic DNA template yielded expected 400-500bp PCR products. Sequences reflecting the *YBX1*sgRNA(1-4)-target specific DNA sequences were PCR amplified and run on a 1% TAE agarose gel. Agarose gel electrophoresis showed that (A) the expected 407bp and 467bp DNA sequences were successfully amplified, representing 400-500bp regions that include the *YBX1*sgRNA1 (293A lane) and *YBX1*sgRNA2 (293B lane) -specific sequences. Likewise, (B) the expected 405bp and 435bp DNA sequences were successfully amplified, representing 400-500bp regions that include the *YBX1*sgRNA3 (lane 293C) and *YBX1*sgRNA4 (lane 293D) -specific sequences were successfully amplified. NTC-A-D reactions for all respective primer pairs showed no contamination. M, Generuler 1kb DNA ladder. NTC, non-template control.

Using the reference genome sequence (NC_000001.11) as control, comparative sequencing analyses enabled identification of the exact mutations causative of the observed knockout

phenotypes. Sequencing results confirmed that successful mutations were introduced by CRISPR/Cas9, especially when compared to reference control. Of note, clone A2 displayed a clean single bp deletion 104nt upstream of the targeted PAM sequence, and was located within exon 1 of the *YBX1* gene (Fig. 3.6A, top sequence read). Sequencing of the B2 clone indicated a clean 10bp deletion that covers intron 4 and exon 5 junction of the *YBX1* gene (Fig. 3.6B, bottom sequence read). In particular an 8bp deletion was observed at the 3'-end of intron 4 and 2bp deleted from the 5'-end of exon 5 of the *YBX1* gene. Sequencing of the C5 clone displayed a more complicated mutation profile suggestive of a heterozygous genotype, where a single nucleotide substitution (A>C) (highlighted bold and underlined in Fig. 3.7) followed immediately by 8bp deletion; this mutation was localised at the 3'-end of exon 6 (Fig. 3.7).

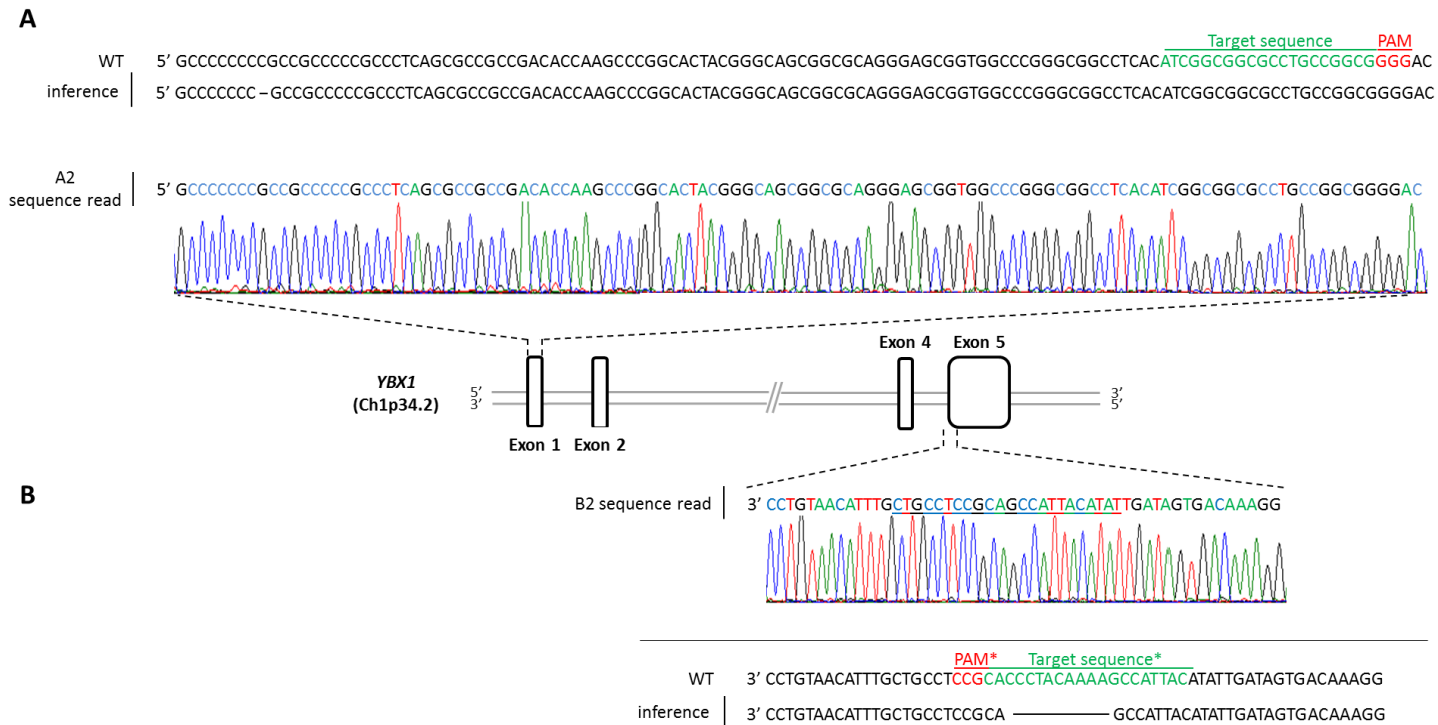


Figure 3.6 Sanger sequencing of YB1 knockout clones A2 and B2 identifies CRISPR/Cas9-mediated mutations. A) Top sequence read reflects Sanger sequencing read and chromatogram of A2 clone showing a single bp deletion (dash line), compared to wildtype reference sequence. **B)** Bottom sequence read reflects Sanger sequencing read and chromatogram for B2 clone, a 10bp deletion (striked line) compared to WT sequence as reference. *, target complementary sequence analysed; PAM, protospacer adjacent motif; WT, wildtype.

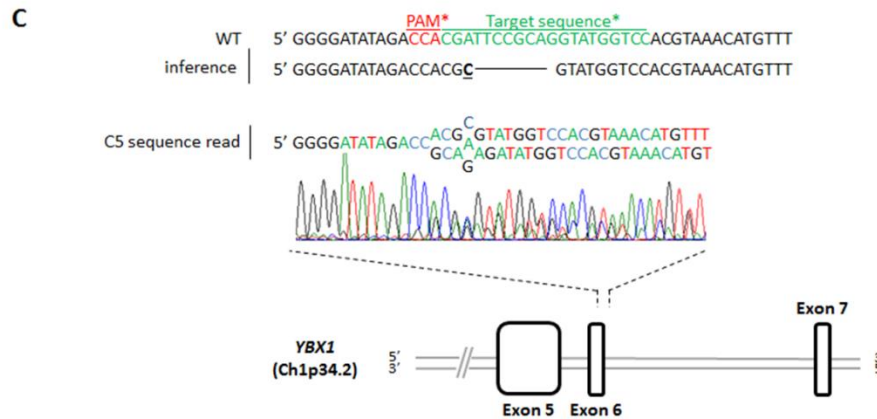


Figure 3.7 Sanger sequencing of YB1 knockout clone C5 identifies CRISPR/Cas9-mediated mutations. Sequence read reflects Sanger sequencing read and chromatogram profile for C5 clone, which displays an A>C substitution (bold and underlined), followed by 8bp deletion (striked line) when compared to wildtype (WT) sequence as reference. *, target complementary sequence analysed; PAM, protospacer adjacent motif; WT, wildtype.

Additionally, 293T parental cells and clones A2, B2 and C5 were kept in culture for an extended period, to examine the stability of the YB1 knockout phenotype in selected clones at 1, 3 and 6 months post-transfection. As a result, sequencing of the equivalent targeted regions from 293T genomic DNA corresponded to the reference sequence (NC_000001.11) at all harvest points (Figs. 3.8A-C, see reference sequence and sequence reads at 1, 3 and 6M). YB1 knockout clones A2 (Fig. 3.8A), B2 (Fig. 3.8B), and C5 (Fig. 3.8C) showed identical sequence profiles as was previously observed (Figs. 3.8A-C, respectively; compare 1M, 3M, and 6M sequence reads to mutated reference sequences). However, further indicative of a heterozygous genotype, the dominant A>C substitution initially observed directly upstream of the 8bp deletion in C5 clone would read either A or C across the timecourse (Fig. 3.8C, see boxed). This was likely a result of both A and C (and even G) nucleotides existing at this particular location across the copies of mutated *YBX1* (Fig. 3.7, refer to C5 sequence chromatogram). The dominant 8bp deletion that followed this A>C substitution was consistent throughout the extended culture period. To conclude, the mutations instilled by the CRISPR/Cas9 genome editing system with the designed gRNAs targeting *YBX1* gene, were considered stable.

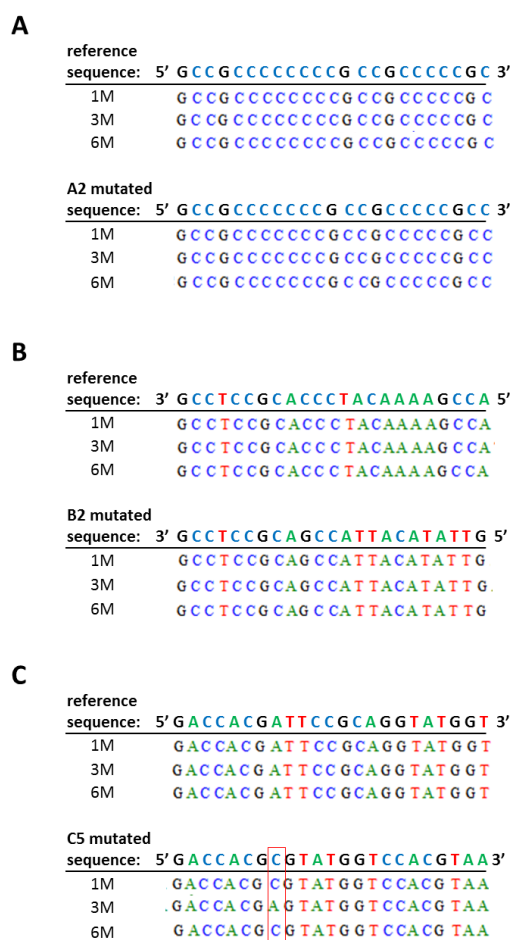


Figure 3.8 Sanger sequencing identifies stable CRISPR/Cas9-mediated mutations in clones A2, B2, and C5 over a 6 month timecourse. 293T and selected clones A2, B2, C5 were cultured and genomic DNA harvested at 1M, 3M and 6M post-transfection for sequencing. **A-C)** 293T control served as reference, and wildtype sequence remained unchanged throughout the 6M period (top reads). The mutations characterised by sequencing were consistently observed for **(A)** A2 – 1bp deletion 104nt upstream of targeted PAM, **(B)** B2 – 10bp deletion, and **(C)** C5 – A>C substitution and 8bp deletion (bottom sequence reads). The exception of A>C substitution was not consistent throughout the entire culture period, with the wildtype A detected at 3M (see boxed). In all instances equivalent sequences sequenced from 293T control demonstrated sequences that correspond to reference sequence (NC_000001.11). 1M, 1 month; 3M, 3 month; 6M, 6 month.

3.2.6 Alternative splicing is a predicted consequence of CRISPR/Cas9 genome editing identified in YB1 knockout clones

Initial screening by Western blotting using different α YB1 indicated that the YB1 knockout genotypes did not disrupt epitope targets for their respective α YB1. However, certain mutations can influence the splicing characteristics of a gene (Lalonde *et al.*, 2017). Therefore, we examined whether or not alternate protein products could result from our CRISPR/Cas9-mediated mutations in selected clones. This was addressed *in silico* using the HSF v3.0 programme (Desmet *et al.*, 2009), which simulated the observed mutations from sequencing, and in turn calculated the potential for changes in splicing events.

A2's 1bp deletion was modelled into the HSF v3.0 as a single cytosine residue deleted at position 40 (Fig. 3.9A). Based on HSF's predictive algorithms no significant splicing motif alterations was reported to be disrupted as a result of the 1bp deletion. The observed mutation probably had no impact on splicing characteristics given the mutation was found well within exon 1. The resulting transcript was nonetheless conceptually translated, and the encoded aa sequence (candidate), given the 1bp deletion, exhibited several premature stop codons when compared to the wildtype YB1 aa sequence (Fig. 3.9B, see asterisks highlighted pink). Multiple sequence analysis between the candidate and wildtype YB1 sequences demonstrated only 20% sequence identity across the length of wildtype YB1 protein (Fig. 3.9C). However, A2's initial most stop codon, which resides at residue 63 (Fig. 3.9C, see first asterisk at position 63 for Query sequence), confers a significantly truncated protein that shares homology comprised of a stretch of 1-15aa to that of wildtype YB1, only.



Figure 3.9 Predictive analysis for alternative splice variants as a result of 1bp deletion in YB1 knockout clone (A2) YBX1 gene. *In silico* analysis using HSF v3.0 of potential and predicted alternative splicing events as a result of the observed 1bp deletion having targeted YBX1 with YBX1sgRNA1, by **A**) modelling the 1bp deletion in YBX1 exon 1 and compared against wildtype splicing events. No measurable splicing alterations around exon 1 were predicted as a result of the 1bp deletion. Figure legend continues next page.

Figure 3.9. Predictive analysis for alternative splice variants as a result of 1bp deletion in YB1 knockout clone (A2) YBX1 gene. Given the 1bp deletion, A2's mutated YBX1 mRNA was input into online protein translator tool, and **(B)** candidate reading frame compared to wildtype YB1 aa sequence. The candidate reading frame was found out of frame of wildtype YB1 reading frame. This is indicated by additional stop codons (denoted as *). **(C)** Multiple sequence analysis between candidate aa sequence (Query) and wildtype YB1 aa sequence (Sbjct) presents minimal (20%) identity.

Clone B2 demonstrated a 10bp deletion across the intron 4:exon 5 splice junction. The fact that Western blot analysis conferred a partial knockout phenotype, despite this significant mutation, led us to postulate a complicated expression profile. The observed mutation was modelled into HSF v3.0 against wildtype YB1 as reference (Fig. 3.10A). Table 3.1 refers to HSF v3.0 consensus value (CV) findings for this modelled deletion compared to wildtype sequence. If a wildtype score is above the CV threshold of 65 and the difference between the wildtype and mutant score (ΔCV) is under -10%, then the wildtype splice site is considered to be broken. On the other hand, when wildtype CV score is below the threshold and the ΔCV is above +10%, then a new splice site is predicted. Therefore, with the B2's mutation modelled, the $\Delta CV\%$ for the wildtype splice site is below -10% (ΔCV , -49.28%) indicating this splice site is presumably disrupted (Table 3.1). On the other hand, alternative splice sites are predicted to be activated (Table 3.1, predictions 2 and 3), where wildtype CV score is below the CV threshold, and $\Delta CV\%$ is above +10% (ΔCV , +270.48 and +254.36%, respectively for predictions 2 and 3). Alternative splice site activation as per prediction 2 conceptually translated a 3aa truncated protein (Fig. 3.10B), corresponding to residues 119-121 of wildtype YB1. In fact, prediction 2 reading frame remains in-frame with wildtype stop codon, which is in contrast to prediction 3 (Fig. 3.10B). The truncation was best demonstrated by multiple sequence alignment analysis, which successfully recognised the truncated protein (Query) with 321/324aa (99% identity) conserved (Fig. 3.10C). This too was in contrast to prediction 3, which demonstrated 45% sequence identity across the entire length of the two sequences (Fig. 3.10D). More importantly, the major aa sequence that is encoded by this alternative splicing product potentially confers a 212aa long protein (corresponding to residues 1-212 before its first stop codon, see first asterisk Fig. 3.10D), of which length showed 64.15% homology (136/212aa), but a reduced 41.9% homology (136/324) when calculated against the entire length of wildtype YB1.

Table 3.1 Predictive assessment for altered splicing motifs after modelling clone B2's mutated sequence into HSF v3.0

Prediction #	Splice site type	Motif	New splice site	WT CV	Mutant CV	$\Delta CV\%$	Impact
1	Acceptor	tggcttttagGG	tggctgcggaggCA	81.49	41.33	-49.28	WT site broken
2	Acceptor	atggcttttagG	atggctgcggagGC	19.95	73.91	+270.48	New site
3	Acceptor	cttttagGGTGC	ctgcggaggcagCA	21.65	76.72	+254.36	New site

Splicing patterns of B2 mutated *YBX1* sequence was compared to wildtype at exon 5 using HSF v3.0. Intron sequences denoted as lower case and exon sequences are denoted as upper case characters. CV, consensus value; ΔCV (%), percent difference between WT CV and Mutant CV; WT, wildtype.

A



B

wildtype YB1

975 nucleotides, 325 amino acids, structure: `sequence AA`

```

1 MSSEAETQQP PAAPPAAPAL SAADTKPGTT GSGAGSGGPG GLTSAAPAGG DKKVIATKVL
61 GTVKWFNVRN GYGFINRNDT KEDVVFHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKGA
121 EAAVNTGPGG VPVQGSKYAA DRNHRYRYPY RRGPPRNYQQ NYQNSSESGEK NEGSESAPEG
181 QAQRRPYRR RRFPPYHRR PYGRRPQYSN PPVQGEVMEG ADNQGAGEQG RPVRQNHRYG
241 YRPRFRFGPP RQRQPRELGN EEDKENQGD E TQQQPPQRR YRRNFNRYRR RPENPKPDG
301 KETKAADPPA ENSSAPEAEQ GGAE
    
```

prediction 2

966 nucleotides, 322 amino acids, structure: `sequence AA`

```

1 MSSEAETQQP PAAPPAAPAL SAADTKPGTT GSGAGSGGPG GLTSAAPAGG DKKVIATKVL
61 GTVKWFNVRN GYGFINRNDT KEDVVFHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKAA
121 NVTGPGGVPV QGSKYAADRN HYRYPYRRR PPRNYQNYQ NSESGEKNEG SESAPEGQAQ
181 QRRPYRRRRF PPRYHRRPY RRPQYSNPPV QGEVMEGADN QGAGEQGRPV RQNHRYGYP
241 RFRFGPPRQ RPRELGNED KENQDETQG QPPQRRYRR NFNRYRRRPE NPKPDGKET
301 KAADPPAENS SAPEAEQGGG E
    
```

prediction 3

962 nucleotides, 320 amino acids, structure: `sequence AA`

```

1 MSSEAETQQP PAAPPAAPAL SAADTKPGTT GSGAGSGGPG GLTSAAPAGG DKKVIATKVL
61 GTVKWFNVRN GYGFINRNDT KEDVVFHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKQM
121 LQVLVVFQFK AVNMQQTVTI IDAIHVGVGL HAITSKITRI VRVGKTRDR RVLKARPNN
181 AGPTAGEGSH LTTGDPMGV DHSIPTLLCR EK*WRVLTTR VQENKVDQ*G RICIGDIDHD
241 SAGALLAKDS LERTAMKKIK KIKEMRPKVS SHLNVGTAAT SITDADAQKT LNHKMAKRQK
301 QPIHQLRIRP LPRLSRAGLS
    
```

Figure 3.10 Predictive analysis for alternative splice variants as a result of 10bp deletion in YB1 knockout clone (B2) YBX1 gene. Figure and legend continues next page.

C

NW Score	Identities	Positives	Gaps
1718	321/324(99%)	321/324(99%)	3/324(0%)
Query 1	MSSEAEQQPPAAPAAPALSAADTKPGTTGSGAGSGGPGGLTSAAPAGGDKKIATKVL		60
Sbjct 1	MSSEAEQQPPAAPAAPALSAADTKPGTTGSGAGSGGPGGLTSAAPAGGDKKIATKVL		60
Query 61	GTVKWFNVRNGYGF INRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEK--		118
Sbjct 61	GTVKWFNVRNGYGF INRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEK		120
Query 119	-AANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSSEGEKNEGSESAPEG		177
Sbjct 121	AANVTGPGGVPVQGSKYAADRNHYRRYPRRRGPPRNYQQNYQNSSEGEKNEGSESAPEG		180
Query 178	QAQQRRPYRRRRFPPYYMRRPYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVQRNMYRG		237
Sbjct 181	QAQQRRPYRRRRFPPYYMRRPYGRRPQYSNPPVQGEVMEGADNQGAGEQGRPVQRNMYRG		240
Query 238	YRPRFRGPPRRQRQPREDGNEEDKENQGDGTGGQPPQRRYRRNFYRRRRPENPKPDG		297
Sbjct 241	YRPRFRGPPRRQRQPREDGNEEDKENQGDGTGGQPPQRRYRRNFYRRRRPENPKPDG		300
Query 298	KETKAADPPAENSSAPEAEQGGAE 321		
Sbjct 301	KETKAADPPAENSSAPEAEQGGAE 324		

D

NW Score	Identities	Positives	Gaps
541	155/345(45%)	182/345(52%)	46/345(13%)
Query 1	MSSEAEQQPPAAPAAPALSAADTKPGTTGSGAGSGGPGGLTSAAPAGGDKKIATKVL		60
Sbjct 1	MSSEAEQQPPAAPAAPALSAADTKPGTTGSGAGSGGPGGLTSAAPAGGDKKIATKVL		60
Query 61	GTVKWFNVRNGYGF INRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEKQM		120
Sbjct 61	GTVKWFNVRNGYGF INRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGETVEFDVVEGEK		120
Query 121	LQVLVVFQKAVNMQQTVTIIDAIVVGVGLHAITSKITRIVRVGKRRDRRVLKARPNN		180
Sbjct 121	-----EAA-----VTGPGGVPVQGSKYAADRNHYR--RYPRRRGPPRNYQQNYQNS		165
Query 181	AGPTAGEGSHLTTGDPHMGV----HSIPTLLCREK*WRVLTTRVQENKVDQ*GRICIGD		236
Sbjct 166	ESGEKNEGSESAPEGQAQQRRPYRRRRFPPYYMRRPYGR---RPQYSNPPVQGEVMEG-		220
Query 237	IDHDSAG--ALLAKDSLERTAMKIKK--IKEMRPKVSSHL---NVG--TAATSITDAD		286
Sbjct 221	ADNQGAGEQGRPVQRNMYRGYRPRFRGPPRRQRQPREDGNEEDKENQGDGTGGQPPQRR		280
Query 287	AQKTLNHKMAKRQKQPIHQ-----LRIRPLRSLRAGLS 320		
Sbjct 281	YRRNFNYRR--RRPENKPDGKETKAADPPAENSSAPEAEQGGAE 324		

Figure 3.10 Predictive analysis for alternative splice variants as a result of 10bp deletion in YB1 knockout clone (B2) YBX1 gene. *In silico* analysis using HSF v3.0 of potential and predicted alternative splicing events as a result of the observed 10bp deletion having had targeted YBX1 with YBX1sgRNA2, by **A**) modelling the 10bp deletion in YBX1 intron 4:exon 5 splice junction and compared against splicing events measured in wildtype sequence. This calculated measurable splicing alterations around intron 4:exon 5 splice junction (see Table 3.1). **B**) Given the 10bp deletion, B2's mutated YBX1 mRNAs for prediction 2 and 3 were input into online protein translator tool, and reading frames compared to wildtype YB1. Prediction 3 reading frame is out of frame of wildtype YB1 reading frame (stop codons are denoted as *). However, prediction 2 reading frame suggests its stop codon is in-frame with wildtype stop codon, and confers a 3aa truncated protein. **C**) Multiple sequence analysis between prediction 2 sequence (Query) and wildtype YB1 aa sequence (Sbjct). **D**) Multiple sequence analysis between prediction 3 sequence (Query) and wildtype YB1 aa sequence (Sbjct).

YB1 knockout clone C5 has shown a heterozygous genotype, as indicated by sequencing, and made calculating potential alternative splicing predictions difficult. This is because the exact mutations that conferred the knockout phenotype for each allele of *YBX1* could not be easily determined by Sanger sequencing. However, sequencing analysis for the C5 clone did show a dominant sequence profile (Figs. 3.6 and 3.7C). This sequence was modelled into HSF v3.0 for the exon 6:intron 6 splice junction (Fig. 3.11A). Table 3.2 refers to HSF v3.0 CV findings for the modelled A>C substitution and 8bp deletion compared to wildtype sequence, and findings suggest that the wildtype splice donor site is potentially broken (Table 3.2, prediction 2; wildtype CV is above HSF threshold and $\Delta CV\%$ is below -10%). However, a potential alternative splice site is predicted to be activated instead as a result of the mutation (wildtype CV below HSF threshold, and $\Delta CV\%$ is above +10%). Prediction 1 extrapolated transcript was conceptually translated and exhibited premature stop codons (Fig. 3.11B, see asterisks), especially when compared to wildtype YB1 sequence. When this aa sequence was compared to full length YB1 by multiple sequence analysis, we see that across the entire lengths of both aa sequences there is shared 79% (259/327aa) homology (Fig. 3.11C). Although, the initial most stop codon occurs for the corresponding residue 253 of YB1, which potentiates in an in-frame reading frame to wildtype YB1 and the expression of a truncated protein with 97.22% (245/252aa) homology. Altogether, potential for alternative splicing was recognised as a consequence of CRISPR/Cas9 genome editing of *YBX1*. This was largely dependent on the localisation of the mutations relative to intron and exon junctions.

Table 3.2 Predictive assessment for altered splicing motifs after modelling clone C5's mutated sequence into HSF v3.0

Prediction #	Splice site type	Motif	New splice site	WT CV	Mutant CV	ΔCV (%)	Predicted impact
1	Donor	CGAttccgc	CGCgtatgg	41.73	73.39	+75.87	New site
2	Donor	CAGgtatgg	CGTatggtc	89.26	33.45	-62.53	WT site broken

Splicing patterns of C5 mutated *YBX1* sequence was compared to wildtype at exon 6:intron 6 using HSF v3.0. Intron sequences denoted as lower case and exon sequences are denoted as upper case characters. CV, consensus value; ΔCV (%), percent difference between WT CV and Mutant CV; WT, wildtype.

A

Reference sequence YBX1 Gene > ENST00000321358 Transcript

```

1 atggtotggt ttgtttgttt ttttaactott ggatttatct ggttgggttt ttttattata ttactgacc agtaggetta atttccattg totttttcag
101 GGTGCTGACA ACCAGGGTGC AGGAGAACAA GGTAGACCAG TGAGGCAGAA TATGTATCGG GGATATAGAC CACGATTCOG CAGgtatggt ccacgtaaac
201 atgtttctat taaaatttcc tcaaacccgt gttaggactc agcaatatca tgcctctcct gcagactcat tttttctgta aca
Total sequence length: 283 nucleotides

```

Mutant sequence In del at position 175 (9 nucleotides deleted, C nucleotides inserted)

```

1 atggtotggt ttgtttgttt ttttaactott ggatttatct ggttgggttt ttttattata ttactgacc agtaggetta atttccattg totttttcag
101 GGTGCTGACA ACCAGGGTGC AGGAGAACAA GGTAGACCAG TGAGGCAGAA TATGTATCGG GGATATAGAC CACG----- --Cgtatggt ccacgtaaac
201 atgtttctat taaaatttcc tcaaacccgt gttaggactc agcaatatca tgcctctcct gcagactcat tttttctgta aca
Total sequence length: 283 nucleotides

```

B

wildtype YB1 975 nucleotides, 325 amino acids, structure: **sequence AA**

```

1 MSSEAEQQP PAAPPAAPAL SAADTKPGTT GSGAGSGGPG GLTSAAPAGG DKKVIATKVL
61 GTVKWFNVRN GYGFINRNDT KEDVFNHQT A IKKNNPRKYL RSVG DGETVE FDVVEGEKGA
121 EAANVTGPGG VPVQGSKYAA DRNHRYRYP RRGPPRNYQQ NYQNSSEGEK NEGSESAPEG
181 QAQRRPYRR RRFPPYMR R PYGRRPQYSN PPVQGEVMEG ADNQGAGEQG RPVRQNMRYG
241 YRPRFRGPP RQRPREDGN EEDKENQGE TQGQQPPQRR YRRNFNRYRR RPENPKPDG
301 KETKAADPPA ENSSAPEAEQ GGAE*

```

prediction 2 967 nucleotides, 322 amino acids, structure: **sequence AA**

```


1 MSSEAEQQP PAAPPAAPAL SAADTKPGTT GSGAGSGGPG GLTSAAPAGG DKKVIATKVL
61 GTVKWFNVRN GYGFINRNDT KEDVFNHQT A IKKNNPRKYL RSVG DGETVE FDVVEGEKGA
121 EAANVTGPGG VPVQGSKYAA DRNHRYRYP RRGPPRNYQQ NYQNSSEGEK NEGSESAPEG
181 QAQRRPYRR RRFPPYMR R PYGRRPQYSN PPVQGEVMEG ADNQGAGEQG RPVRQNMRYG
241 YRPRGSSPK TA*RGRO*RR *RKSRR*DPR SAATSTSVPP QLQLPTQTPR KP*TTRWQRD
301 KSSRSTS*EF VRSRG*AGRG *V

```

C

	NW Score	Identities	Positives	Gaps
	1308	259/327(79%)	267/327(81%)	8/327(2%)
Query 1		MSSEAEQQP PAAPPAAPAL SAADTKPGTTGSGAGSGGPGGLTSAAPAGGDKKVIATKVL		60
Sbjct 1		MSSEAEQQP PAAPPAAPAL SAADTKPGTTGSGAGSGGPGGLTSAAPAGGDKKVIATKVL		60
Query 61		GTVKWFNVRNGYGF INRNDTKEDVFNHQT AIKKNNPRKYL RSVG DGETVEFDVVEGEKGA		120
Sbjct 61		GTVKWFNVRNGYGF INRNDTKEDVFNHQT AIKKNNPRKYL RSVG DGETVEFDVVEGEKGA		120
Query 121		EAANVTGPGGVPVQGSKYAADRNHYRYP RRRGPPRNYQQ NYQNSSEGEK NEGSESAPEG		180
Sbjct 121		EAANVTGPGGVPVQGSKYAADRNHYRYP RRRGPPRNYQQ NYQNSSEGEK NEGSESAPEG		180
Query 181		QAQRRPYRRRRFPYMRPYGRRPQYSNPPVQGEVMEGADNQAGEQGRPVQRNMRYG		240
Sbjct 181		QAQRRPYRRRRFPYMRPYGRRPQYSNPPVQGEVMEGADNQAGEQGRPVQRNMRYG		240
Query 241		YRPRGSSPKTA*RGRO*RR*RKSRR*DPRSAATSTSVPPQLQLPTQ---TPRKP*TTRW		297
Sbjct 241		YRPRFRGPP---RQRPREDGNEEDKENQGEDETQGQQPPQRRYRRNFNRYRRRRPENPKP		297
Query 298		QRDKSSRSTS*EFVRSRG*AGR--G*V 322		
Sbjct 298		Q K +++ S G 324		

Figure 3.11 Predictive analysis for alternative splice variants as a result of A>C substitution and 8bp deletion in YB1 knockout clone (C5) YBX1 gene. The A>C substitution and 8bp deletion in YBX1 exon 6:intron 6 splice junction was modelled into HSF v3.0, and compared against splicing events measured in wildtype sequence. HSF calculated measurable splicing alterations around exon 6:intron 6 splice junction (see Table 3.2). Figure legend continues next page.

Figure 3.11. Predictive analysis for alternative splice variants as a result of A>C substitution and 8bp deletion in YB1 knockout clone (C5) YBX1 gene. B) Given the A>C substitution and 8bp deletion, C5's mutated YBX1 mRNA for prediction 2 was inputted into an online protein translator tool, and reading frame compared to wildtype YB1 aa sequence, demonstrating the prediction 1 reading frame is out of frame of wildtype YB1 reading frame (stop codons are denoted as ). **C)** Multiple sequence analysis between prediction 2aa sequence (Query) and wildtype YB1 aa sequence (Sbjct).

3.2.7 Surveyor[®] mutation screening identifies heterozygous and homozygous genotypes for YB1 knockout clones

Surveyor[®] mutation detection assay is typically used to estimate a rate of genome editing post-transfection or treatment with CRISPR/Cas9 (Qiu *et al.*, 2004). However, single cell clones were generated in this study, and the use of the mutation screening technique was adjusted to examine the hetero- or homo- zygous genotypes of YB1 knockout clones. Instead, Surveyor[®] mutation detection assay was used to further characterise of YB1 knockout cell lines was made possible through Surveyor[®] mutation detection assays. This used a sensitive endonuclease (nuclease S) to cleave mismatches present in duplex DNA (Vuillot *et al.*, 2015). Therefore, CRISPR/Cas9-targeted sequences could be PCR amplified, denatured and hybridised randomly to form homoduplexes and heteroduplexes (depending on the genotype nature of cloned cell populations), and then subjected to nuclease S digestion to digest all possible mismatches (including single nucleotide mismatches) and visualisation.

Given that small deletions were observed for clones A2, B2 and C5 by sequencing, little difference was expected to be observed in the resolving nature of each PCR product when compared to the equivalent, amplified products from 293T control (Figs. 3.12A-C). Specifically, approximately 407, 467 and 405bp PCR amplicons expected for the corresponding YBX1sgRNA1-3 CRISPR/Cas9-targeted genomic regions (clones A2, B2, and C5, respectively), were identified between 293T, A2, and spiked samples (Figs. 3.12A-C). 293T control was confirmed as homozygous for each respective CRISPR/Cas9-targeted region of the YBX1 gene; because the corresponding targeted regions showed no nuclease S-mediated digestion products (Figs. 3.12D-F, see 293T control lanes). Only the expectant PCR products of approximately 407, 467, and 405bp, corresponding to the equivalent CRISPR/Cas9 targeted regions identified for exon 1, exon 5, and intron 6, respectively, was evident. Therefore, clone A2 also exhibited a homozygous genotype for the observed 1bp mutation, because nuclease S digestion products were not observed below the approximate 407bp homoduplex band (Fig. 3.12D, see A2 lane). Expectant fragments of approximately 194 and 183bp occurred post-nuclease S treatment only with regards to heterozygous simulated reactions (Fig. 3.12D). These involved spiking in A2 genomic DNA into 293T genomic DNA as template to simulate a

heterozygous genotype after heteroduplex formation. Additionally, this setup exhibited a dose-dependent characteristic, where an increasing ratio of A2 genomic DNA spiked into 293T control genomic DNA template correlated with more prominent digestion bands, compared to reactions involving a lower ratio of A2 genomic DNA spike. Similarly, a similar homozygous genotype was also observed for B2 clone (Fig. 3.12E). Expectant digestion products of approximately 302 and 165bp were only evident post-nuclease S treatment and concerning heterozygous-simulated reactions with B2 genomic DNA spiked into 293T genomic DNA as template. In contrast, clone C5 produced expected digestion bands post-nuclease S digestion even without spiking in C5 genomic DNA into 293T genomic DNA for PCR template (Fig. 3.12F). Expected bands of approximately 280 and 125bp were evident for C5 lane. In all cases, prominent digestion bands for A2, B2, and C5 were observed at higher ratios of YB1 knockout clone:293T genomic DNA as template to simulate the heterozygous genotype control. Overall, Surveyor® mutation screenings suggested that both homozygous and heterozygous genotypes were observed after CRISPR/Cas9-mediated genome editing, despite subjecting transfected cells for single cell cloning.

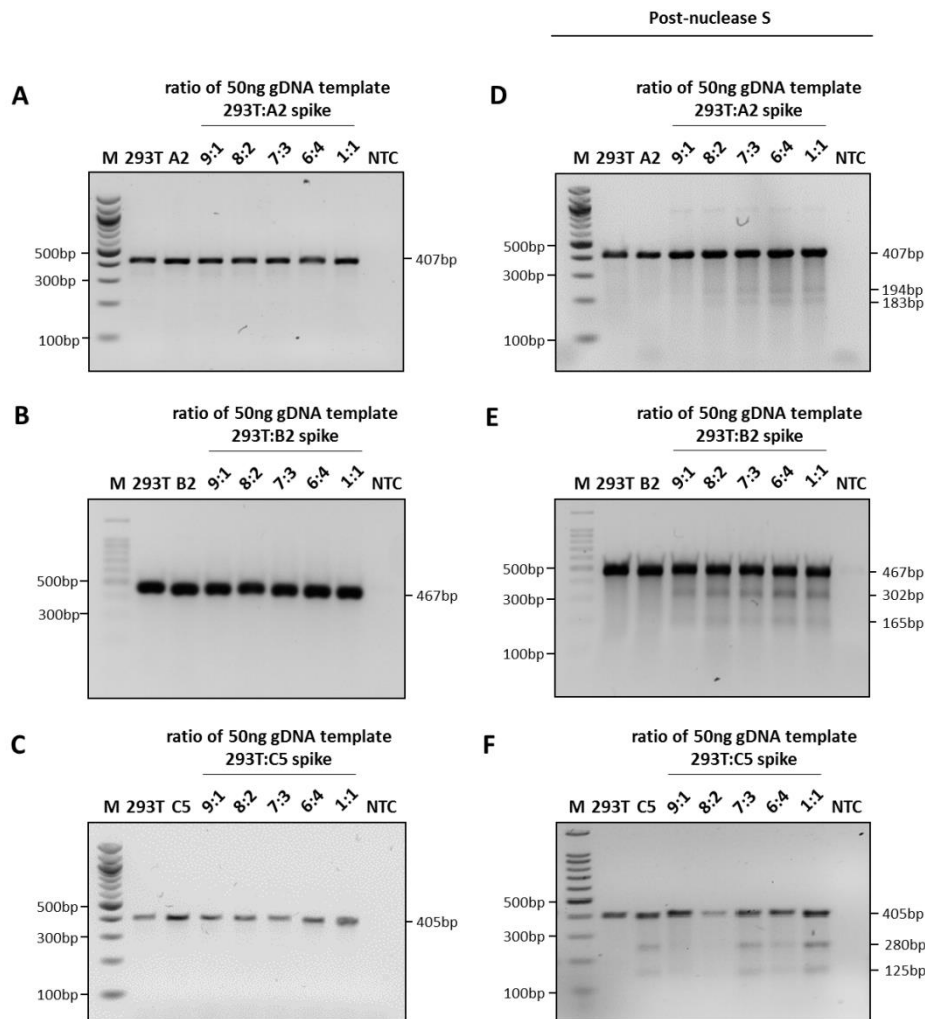


Figure 3.12 Surveyor[®] mutation screening identifies homozygous and heterozygous genotypes by CRISPR/Cas9-genome editing of *YBX1*. Mutated sequences previously observed by Sanger sequencing in selected clones A2, B2, and C5, and the corresponding regions in 293T control were PCR amplified and subjected to Surveyor[®] mutation screening. In addition to this, PCR were performed to simulate heterozygosity control by spiking 293T genomic DNA with *YB1* knockout clone (A2, B2, or C5) genomic DNA at the indicated ratios (50ng total DNA template). **A-C)** Expectant PCR amplicons of approximate size was evidenced by 2% TBE agarose gel electrophoresis for 293T and corresponding regions in selected *YB1* knockout clones (407, 467 and 405bp PCR amplicons for A2, B2, and C5 mutated regions, respectively), respectively. PCR samples were then subjected to nuclease S-digestion and (D-F) digestion fragments identified by 2% TBE agarose gel electrophoresis. Digestion bands were only evident for heterozygous-control samples for (D) A2, (E) B2, and (F) C5, with C5 exhibiting a heterozygous genotype given digestion bands were evident in C5 only lane, post-nuclease S treatment. M, 100bp DNA ladder; NTC, non-template control.

3.2.8 High resolution melting (HRM) curve analysis distinguishes wildtype alleles from mutated

To explore whether wildtype *YBX1* alleles were responsible for the heterozygosity profile shown for clone C5, more sensitive genotyping was performed in the form of HRM curve analyses. This was because the sensitivity of the Surveyor[®] mutation assays and sequencing analysis were unable to absolutely distinguish heterozygosity in the context of the wildtype *YBX1* allele. Essentially, discrete differences in nucleotide composition can be distinguished by

the differences in melting dynamics by HRM curve analysis, when compared to the equivalent region from a wildtype reference (Wittwer *et al.*, 2003).

Taking clone C5, the CRISPR/Cas9-targeted DNA region that was amplified (referred to as KOC) was assessed in parallel with the corresponding DNA region from 293T control (referred to as 293C, and functioned as a baseline reference). Dissociation curve analysis (Fig. 3.13A) revealed three distinct dissociation peaks (referred to as KOC1-3, black line) separately observed for C5 mutated region in quadruplicate compared to 293T control (red line). KOC1-3 dissociation peaks showed average dissociation temperatures of $76.65 \pm 0.02^\circ\text{C}$, $75.56 \pm 0.01^\circ\text{C}$, and $74.75 \pm 0.01^\circ\text{C}$, respectively, and relatively distinct from 293C control (Figs. 3.13A and B). C5's dissociation peaks were all significantly different to 293C's dissociation peak average of $77.66 \pm 0.14^\circ\text{C}$ (Fig. 3.13B, annotated peak WT), derived from 293T as control ($n = 4$; $P < 0.01$ and $P < 0.001$). Therefore dissociation peaks KOC1, KOC2 and KOC3 showed 1.01°C ($P < 0.01$), 2.10°C ($P < 0.001$), and 2.91°C ($P < 0.001$) difference in dissociation peak temperature compared to 293C, respectively (Fig. 3.13B). Expected HRM-PCR amplicons of approximate size was identified by 2% TBE agarose gel electrophoresis indicated that the 293C and KOC sequences were successfully amplified as 93bp PCR products, and non-template control reaction yielded only primer dimers (Fig. 13.3D), that did not deter above analysis.

Melting curves for clone C5's KOC and 293T's 293C amplicons suggests that significant changes in melt curve properties were present in the mutated region compared to 293C control (Fig. 3.13E). The KOC melt region (black line) was distinct from 293C wildtype sequence (red line), and that the CRISPR/Cas9-mediated genome editing within the targeted region overall reduced the melting temperature properties from wildtype sequence. Again, HRM reactions run on a 2% agarose gel (Fig. 3.13D) imply that the melt curves are reflective of the expected DNA sequence targets (approximately 93bp), and further distinct primer dimers from non-template control reactions. Furthermore, difference graph analysis of normalised melt curves, after applying 293C's genotype as homozygous, detected variation when compared to KOC (Fig. 3.13E). Therefore, normalised melt curve analysis indicated that i) KOC showed variation such that the analysed DNA sequence exhibited no homozygosity for mutated *YBX1* copies, and ii) KOC DNA sequence was distinct from the assigned reference control. Furthermore, difference graphs also indicated that 293C HRM profiles were similar between each technical replicate, except for one particular replicate, which was detected as showing 'variation', but with 85.05% confidence value.

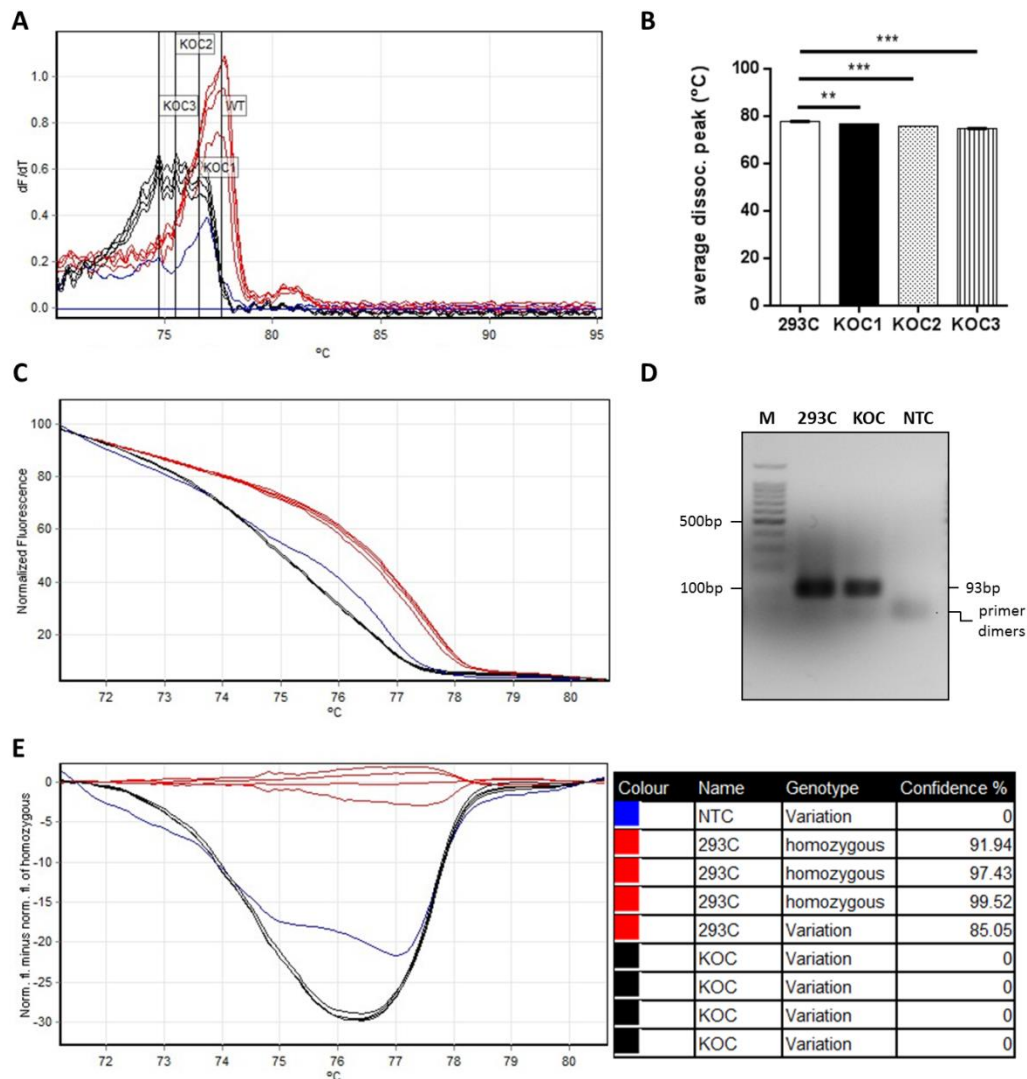


Figure 3.13 HRM curve analysis of YB1 knockout clone C5 distinguishes a heterozygous genotype, without wildtype YBX1 allele. HRM curve analysis demonstrated that the mutated sequences in YB1 knockout clone C5 were distinct from the wildtype YBX1 sequence, and further gives evidence of a heterozygous genotype. **A)** Dissociation curve analysis displays multiple peak dissociation temperatures (indicative of a heterozygous mutation profile) of C5 mutated region (KOC1-3) compared a single peak dissociation temperature to the equivalent region in 293T control (293C). Peak dissociation temperatures were annotated as WT for 293C (red line), and KOC1-3 for each KOC-derived peak (black line). **B)** Average dissociation temperatures plotted for 293C (solid white bar, 77.66°C) and KOC1-3 (solid black bar, 76.65°C; dotted bar, 75.56°C; and vertical lines bar, 74.75°C, respectively). Error bars reflect \pm SD from mean, $n = 4$; **, $P < 0.01$; ***, $P < 0.001$. **C)** Normalised melt curve profile between 293C (red line) and KOC (black line), portraying the impact C5's A>C substitution and 8bp deletion has on the curvature of melt curve from 293T (293C) control. **D)** 2% TBE agarose gel electrophoresis of HRM reactions showing expected PCR products for 239C and KOC (approximately 93bp), with primer dimer formation evident in NTC lane. **E)** Difference graph between 293C (red line) and KOC (black line), and autocalled inference of samples' genotype when compared against 293C as homozygous control. Confidence threshold was set to 90%. NTC, non-template control; M, 100bp DNA ladder; WT, wildtype.

In contrast, dissociation peaks and melt curve properties were comparable between 293T and clone C5 for either sequences that immediately flank the amplicon harbouring the observed mutation (Figs. 3.14 and 3.15, for 5' and 3' -flanking controls respectively). These regions

functioned as controls for CRISPR/Cas9-mediated mutations, as these flanking sequences were not targeted by Cas9:sgRNA, and were observed intact by sequencing analyses. Therefore 5'-flanking sequences were referred to as 5'293C or 5'KOC, and 3'-flanking sequences were referred to as 3'293C or 3'KOC. When examining the 5'-flanking sequences, an average dissociation peak temperature of $82.04 \pm 0.01^\circ\text{C}$ and $81.97 \pm 0.04^\circ\text{C}$ was detected for 5'293C (red line) and 5'KOC (black line), respectively (Figs. 3.14A and B); and the calculated difference of 0.07°C showed no significant difference ($n = 4$, $P > 0.05$). The intended HRM-PCR products were correctly amplified (approximately 95bp by 2% TBE agarose electrophoresis), and suggested that the dissociation peaks correlated with the 5'293C and 5'KOC control sequences, and distinct from primer dimers (Fig. 3.14D). Additionally, normalised melt curves presented with similar curvatures between 5'293C (red line) and 5'KOC (black line) (Fig. 3.14C). Further scrutiny between 5'293C and 5'KOC melt curves was made by difference plots (Fig. 3.14E), which identify 5'293C as homozygous and similar between its replicates. Interestingly, however, 3/4 5'KOC replicates showed variation when compared to the homozygous genotype of the 5'293C reference control, albeit with 70-86% confidence below the 90% cut-off.

The 3'-flanking sequence was also analysed by dissociation curve analysis, and showed average dissociation peaks of $78.53 \pm 0.43^\circ\text{C}$ and $78.51 \pm 0.03^\circ\text{C}$ for 3'293C (red line) and 3'KOC (black line) for downstream flanking DNA sequence, respectively. The difference of 0.02°C in dissociation temperature also showed no significant difference ($n = 4$, $P > 0.05$) (Figs. 3.15A and B). Intended HRM-PCR products (approximately 104bp by 2% TBE agarose gel electrophoresis) were identified, suggesting that the dissociation peaks measured correlate with the 3'293C and 3'KOC control sequences, and distinct from primer dimers (Fig. 3.15D). Melt curve profiles were also comparable between 3'293C (black line) and 3'KOC (red line) with regards to curvature of normalised melt curves (Fig. 3.15C). Further to this, difference plots were generated with 3'293C set with a homozygous genotype as control (Fig. 3.15E), and 3'KOC regions were successfully recognised as homozygous in genotype with confidence values all exceeding the threshold of 90% for 2/4 replicates. Altogether, HRM analysis provided the additional required sensitivity to confidently characterise a heterozygous genotype for a particular mutation, and distinguish from wildtype, simultaneously.

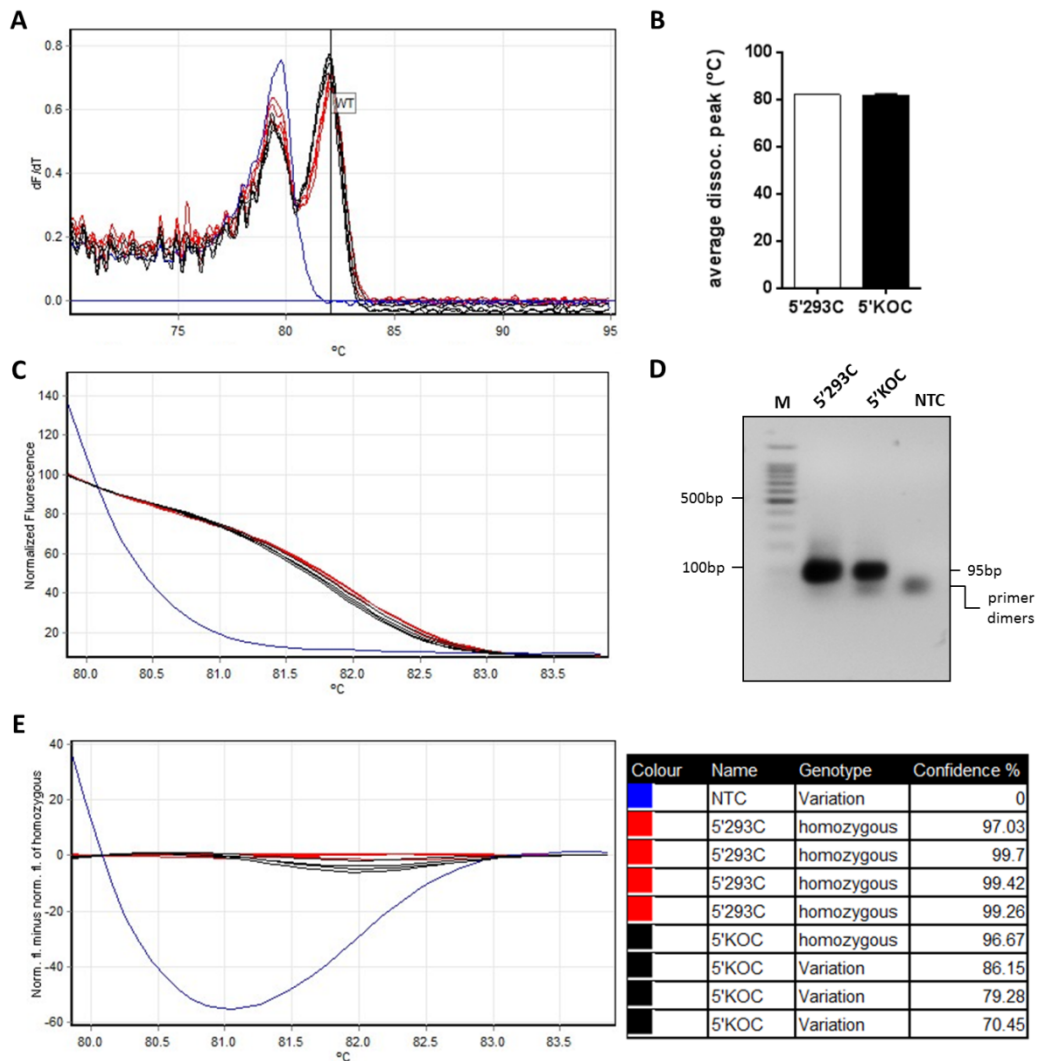


Figure 3.14 HRM curve analysis of sequences upstream of the mutated YBX1 sequence identifies as comparable to 293T control. A) Dissociation curve analysis displays peak dissociation temperatures of 5'KOC sequence compared to the equivalent region in 293T control (5'293C). Peak dissociation temperatures are annotated as WT for both 5'293C (red line), and 5'KOC peaks (black line). **B)** Average dissociation temperatures plotted for 5'293C (solid white bar, 82.04°C) and 5'KOC (solid black bar, 81.97°C), respectively. Error bars reflect \pm SD of mean, $n = 4$. **C)** Normalised melt curve profile between 5'293C (red line) and 5'KOC (black line). **D)** 2% TBE agarose gel electrophoresis of HRM reactions showing expected PCR products (approximately 95bp) for 5'293C and 5'KOC, with primer dimer formation evident in NTC lane. **E)** Difference graph between 5'293C (red line) and 5'KOC (black line), and autocalled inference of sample genotype, based on difference graph analysis and setting 5'293C as homozygous control, with confidence threshold set to 90%. NTC, non-template control; M, 100bp DNA ladder; WT, wildtype.

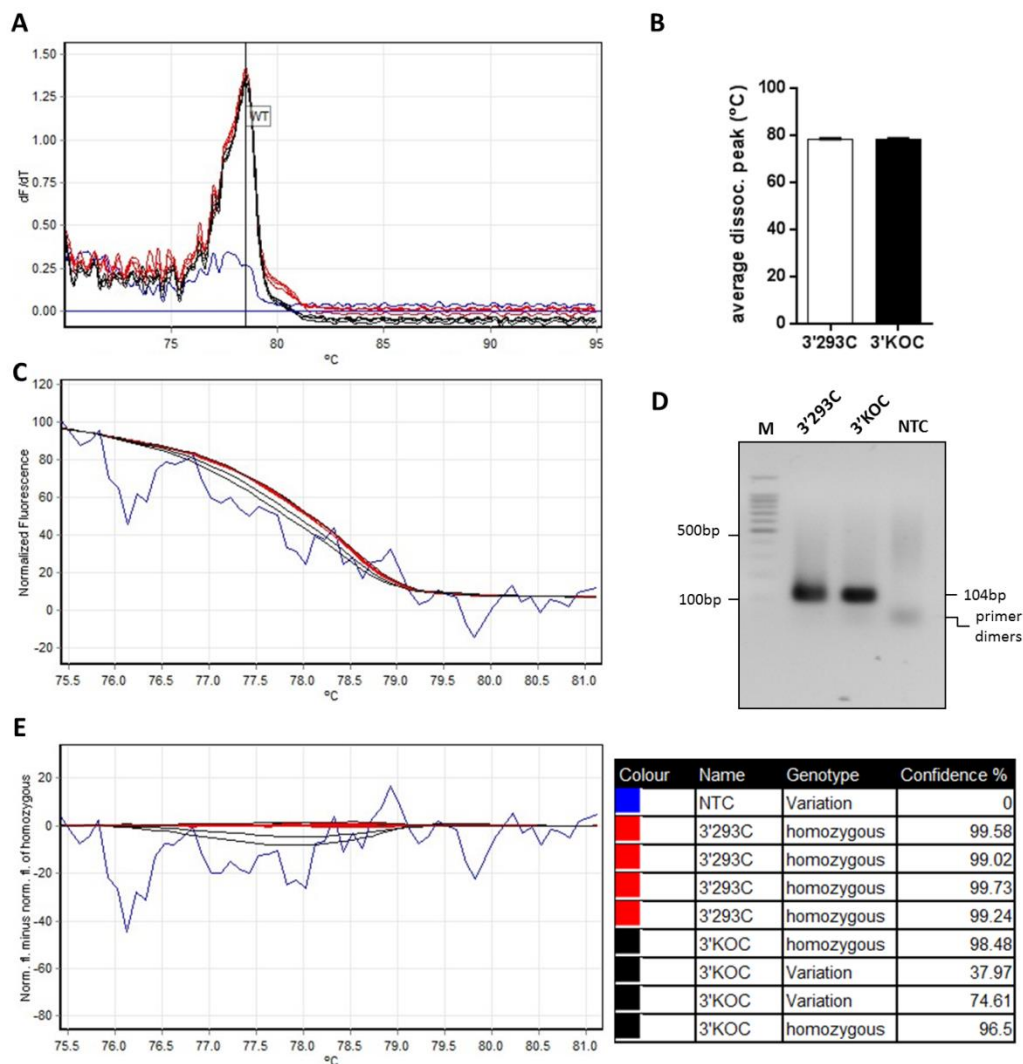


Figure 3.15 HRM curve analysis of sequences downstream of the mutated YBX1 sequence identifies as comparable to 293T control. A) Dissociation curve analysis displays peak dissociation temperatures of 3'KOC sequence compared to the equivalent region in 293T control (3'293C). Peak dissociation temperatures are annotated as WT for both 3'293C (red line), and 3'KOC peaks (black line). **B)** Average dissociation temperatures plotted for 3'293C (solid white bar, 78.51°C) and 3'KOC (solid black bar, 78.53°C), respectively. **C)** Normalised melt curve profile between 3'293C (red line) and 3'KOC (black line). **D)** 2% TBE agarose gel electrophoresis of HRM reactions showing expected PCR products for 3'293C and 3'KOC, with primer dimer formation evident in NTC lane. **E)** Difference graph between 3'293C (red line) and 3'KOC (black line), and autocalled inference of sample genotype, based on difference graph analysis and setting 3'293C as homozygous control, with confidence threshold set to 90%. Error bars reflect \pm SD of mean, $n = 4$. NTC, non-template control; M, 100bp DNA ladder; WT, wildtype.

3.2.9 Stable Cas9 expression was not evident in YB1 knockout clones

YB1 knockout cell lines: A2, B2 and C5, were further screened for stable Cas9 expression over a 3 or 12 months timecourse, alongside bulk transfected populations (harvested at 72h and at passage 2 post-transfection with CRISPR/Cas9-nuclease vectors). Non-transfected, parental 293T cells served as a negative control for Cas9 expression. The safety profile of the newly

generated cell lines was examined in context of stable Cas9 expression – a bacterium derived endonuclease. Additionally, the long-term stability of the YB1 knockout phenotype was examined to ensure that the novel cell lines maintained the desired phenotype.

Initially, non-transfected 293T cells cultured upto 12 months were harvested and subjected to Western blotting to examine the expression profile of YB1 protein long-term (Fig. 3.16A). Given that equal loading was determined by GAPDH, YB1 levels were considered comparable between lysates derived from non-transfected 293T harvested at 72h, 1month, and 12 months in culture (Fig. 3.16A). Therefore, it was considered that long-term culture of 293T cells did not affect the expression of endogenous YB1. Non-transfected 293T cells also demonstrated the lack of Cas9 protein expression, including in subsequent Western blot analyses (Fig. 3.16B-D), as expected for negative transfection controls. Cas9 expression was, however, detectable only in bulk transfected 293T cells with pCRISPR-*YBX1*sgRNA1-3 for *YBX1* targeting via exon 1, 5 and intron 6, respectively (Figs. 3.16B-D, respectively). Especially, Cas9 expression levels were markedly expressed 72h post-transfection in all cases of bulk transfected 293T populations, but significantly diminished by passage 2 for each pCRISPR-*YBX1*sgRNA1-3 (Figs. 3.16B-D, respectively). The levels of Cas9 protein expression was found completely undetectable in all clonal populations cultured from 1 month onwards up to 12 months for the A2 YB1 knockout cell line (Fig. 3.16B). Similar was observed for B2 and C5 clones from passage 5 onwards upto 3 months in culture (Figs. 3.16C and D). Additionally, given normalised loads as per GAPDH control, the YB1 knockout phenotype was maintained and considered stable in A2 and C5 clones up to 12 months and 3 months in culture, respectively (Figs. 3.16B and D); whereas clone B2 maintained a partial knockout phenotype upto 3 months in culture (Fig. 3.16C). Otherwise, YB1 levels were only detectable by Western blotting for non-transfected 293T control, and transfected 293T cells prior to single cell cloning up to passage 2 post-transfection.

Therefore, the phenotypes first identified by Western blotting of the selected clones (Fig. 3.4) were also considered stable up to 3 or 12 months in culture having had used CRISPR/Cas9 genome editing to target the *YBX1* gene. Additionally, the transient nature of Cas9 expression by pCRISPR plasmid transfection is demonstrated by all three pCRISPR-*YBX1*sgRNA1-3 vectors used in the current study. Rather, stable Cas9 expression, upto 12 months in culture in select YB1 knockout clones, was not evident by the limits of detection by Western blotting.

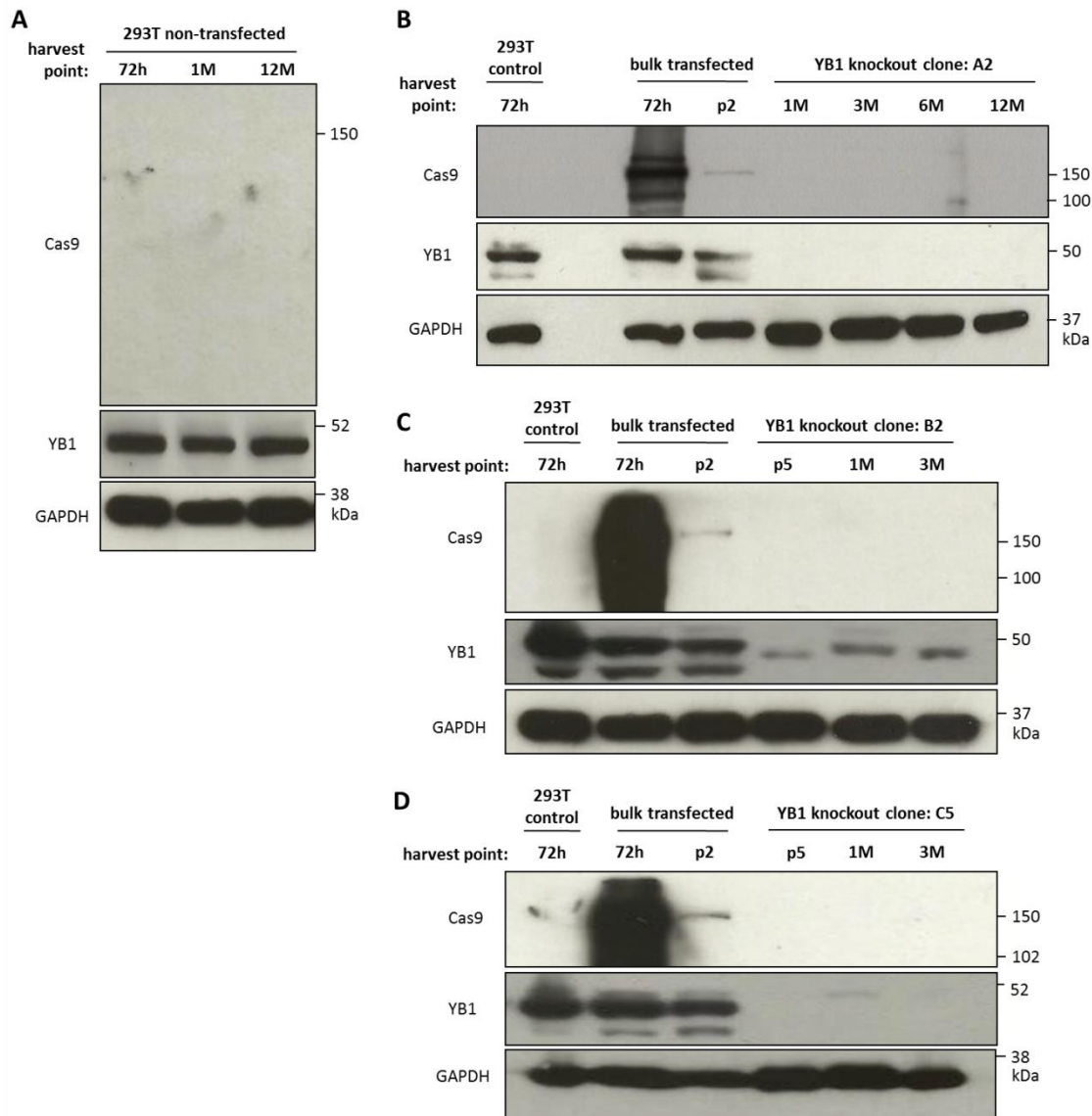


Figure 3.16 Western blot analyses indicate Cas9 expression is transiently expressed after pCRISPR-*YBX1*sgRNA1-3 transfections. Non-transfected and bulk transfected 293T and selected YB1 knockout clones were maintained and harvested at the indicated time-points upto 3 or 12 months in culture, and lysates subjected to Western blotting to assess stable Cas9 expression (top blots) and YB1 phenotypes (middle blots). **A)** Non-transfected 293T cells served as negative controls for Cas9 expression, but YB1 expression remained unchanged despite the extended culture period. Western blot analysis suggests that Cas9 expression was transient (exhibited only by bulk transfected 293T 72h and passage 2 post-transfection), and YB1 knockout phenotypes were maintained for **(B)** A2, **(C)** B2, and **(D)** C5 clones, after CRISPR/Cas9 targeting of *YBX1* using pCRISPR-*YBX1*sgRNA1-3, respectively. GAPDH (bottom blots) was used as a loading control. p2, passage 2 post-transfection; p5, passage 5 post-transfection; 1M, 1 month; 3M, 3 months; 6M, 6 months; 12M, 12 months.

3.2.10 Off-target effects in YB1 knockout clones by Western Blotting

Additional safety features were examined in the selected YB1 knockout clones, and involved measuring for off-target effects by CRISPR/Cas9-mediated genome editing, principally by Western blotting. Genome editing involves permanent alterations in the genome, and target

specificity of Cas9 is dependent on homology driven pairing between the gRNA and target sequence (Lim *et al.*, 2016). However, Cas9 can direct its endonuclease function off-target due to the flexibility allowed in the base-pairing and complementarity between gRNA and target DNA sequence (Mali *et al.*, 2013b). We therefore assessed any change in off-target protein expression when compared to non-transfected 293T as baseline control. Potential off-target sites were identified *in silico* by BLAST analysis of *YBX1*sgRNA1-3 designs. The top five hits for each are listed below in Tables 3.3, 3.4, and 3.5, respectively.

Despite each designed *YBX1*sgRNA showed 100% homology against the intended *YBX1* gene, *in silico* screening identified potential off-target sites for *YBX1*sgRNA1 (Table 3.3), which included: 5'-side of exotosin-like glycosyltransferase 1 (*EXTL1*; 75% homology), pantothenate kinase 6 (*PANK4*; 70% homology), 5'-side of SRY-Box 13 (*SOX13*; 70% homology), serine incorporator 2 (*SERINC2*; 70% homology), and 5'-side of stromal cell derived factor 4 (*SDF4*; 65% homology) genes. Potential off-target sites for *YBX1*sgRNA2 (Table 3.4) instead included: 5'-side of regulator complex protein (*LAMTOR5*; 85% homology), cAMP-dependent kinase catalytic subunit β isoform (*PRKACB*, 75% homology), tudor domain-containing protein 10 isoform B (*TDR10*; 75% homology), 5'-side of olfactory receptor 2G3 (*OR2G3*; 75% homology), and 5'-side of olfactory receptor 10J3 (*OR10J3*; 70% homology). Potential off-target sites for *YBX1*sgRNA3 (Table 3.5), however, included: 5'-end of leptin gene-related protein isoform 2 (*LEPROT*; 65% homology), collagen α -1(XXIV) chain precursor (*COL24A1*; 65% homology), ubiquitin carboxyl hydrolase isozyme L5 isoform 2 (*UCHL5*; 65% homology), 5'-side of β -1,3-galactosyltransferase 2 (*B3GALT2*; 65% homology), and finally, piggyback transposable element derived protein 5 (*PGBD5*; 65% homology).

It is appreciated that a number of these potential off-target sites were found in the untranslated regions or further upstream/downstream to the associated genes' coding sequence. Off-target sites that matched this parameter were annotated with an asterisk to the associated gene name (Tables 3.3-5). For example, the off-target site for *YBX1*sgRNA1 in association with *EXTL1* was identified 40-55bp downstream of the gene's stop TAG codon. Analyses also involved identifying 5'-NGG PAM sequences associated with these potential off-target sites; and no 5'NGG PAMs were recognised following the off-target sequences. The only exception to this was the off-target site identified downstream of the *EXTL1* gene, of which off-target sequence was found to precede a 5'-NGG-3' PAM sequence (Table 3.3).

Table 3.3 Potential off-target sites identified for gRNA1 targeting of *YBX1* gene

number	gRNA1 vs Subject (5' > 3') (coverage and sequence)				Homology (%)	PAM sequence present?	Gene identified	Related protein function		
1	gRNA1: 6	CGGCGCCTGCCGGCG	20	Subject: 6035402	CGGCGCCTGCCGGCG	26035416	75	Yes	<i>EXTL1*</i>	Chain polymerisation of heparan sulphate and heparin (Kim <i>et al.</i> , 2003).
2	gRNA1: 5	GCGGCGCCTGCCGG	18	Subject: 526229	GCGGCGCCTGCCGG	2526242	70	No	<i>PANK4</i>	Regulatory enzyme for biosynthesis of coenzyme A (CoA) (Halvorsen and Skrede, 1982).
3	gRNA1: 3	CGGCGGCGCCTGCC	16	Subject: 204129200	CGGCGGCGCCTGCC	204129213	70	No	<i>SOX13*</i>	Transcription factor involved in regulating cell fate and embryonic development (Baroti <i>et al.</i> , 2016).
4	gRNA1: 6	CGGCGCCTGCCGGC	19	Subject: 31413520	CGGCGCCTGCCGGC	31413507	70	No	<i>SERINC2</i>	Transmembrane protein involved in incorporation of serine into phosphatidylserine and sphingolipid (Inuzuka <i>et al.</i> , 2005).
5	gRNA1: 3	CGGCGGCGCCTGC	15	Subject: 1232222	CGGCGGCGCCTGC	1232234	65	No	<i>SDF4*</i>	Ca ²⁺ -binding protein localised to Golgi apparatus (Scherer <i>et al.</i> , 1996)

* denotes potential off-target sites that were found in UTR sequences or further upstream/downstream of the associated gene's coding sequence.

Table 3.4 Potential off-target sites identified for gRNA2 targeting of *YBX1* gene

number	gRNA2 vs Subject (5' > 3') (coverage and sequence)				Homology (%)	PAM sequence present?	Gene(s) identified	Related protein function
1	gRNA1: 1	GTAATGGCTTTTGTAGG	17		85	No	<i>LAMTOR5*</i>	Involved in sensing aa and activation of mTORC1 (Bar-Peled <i>et al.</i> , 2012).
	Subject: 110418735	GTAATGGCTTTTGTAGG	110418751					Catalytic subunit of cAMP-dependent protein kinase, to mediate signalling for multiple cell processes (Wu <i>et al.</i> , 2002).
2	gRNA1: 5	TGGCTTTTGTAGGGT	19		75	No	<i>PRKACB</i>	Component of cytoplasmic RNA granules and regulates post-transcriptional activities of specific mRNAs (inferred from electronic annotation).
	Subject: 84223967	TGGCTTTTGTAGGGT	84223981					Encodes olfactory receptor (inferred from electronic annotation).
3	gRNA1: 5	TGGCTTTTGTAGGGT	19		75	No	<i>TDRD10</i>	Encodes olfactory receptor (inferred from electronic annotation).
	Subject: 154524687	TGGCTTTTGTAGGGT	154524701					Encodes olfactory receptor (inferred from electronic annotation).
4	gRNA1: 1	GTAATGGCTTTTGTAGG	15		75	No	<i>OR2G3*</i>	Encodes olfactory receptor (inferred from electronic annotation).
	Subject: 247613456	GTAATGGCTTTTGTAGG	247613442					Encodes olfactory receptor (inferred from electronic annotation).
5	gRNA1: 5	TGGCTTTTGTAGGG	18		70	No	<i>OR10J3*</i>	Encodes olfactory receptor (inferred from electronic annotation).
	Subject: 159422240	TGGCTTTTGTAGGG	159422253					Encodes olfactory receptor (inferred from electronic annotation).

* denotes potential off-target sites that were found in UTR sequences or further upstream/downstream of the associated gene's coding sequence.

Table 3.5 Potential off-target sites identified for gRNA3 targeting of *YBX1* gene

number	gRNA3 vs Subject (5' > 3') (coverage and sequence)				Homology (%)	PAM sequence present?	Gene identified	Related protein function
1	gRNA1:	3	ACCATACCTGCGG	15	65	No	<i>LEPROT*</i>	Negatively regulates leptin receptor expression to cell surface (Couturier <i>et al.</i> , 2007).
	Subject:	65526447	ACCATACCTGCGG	65526435				
2	gRNA1:	1	GGACCATACCTGC	13	65	No	<i>COL24A1</i>	Potentially helps regulate type I collagen fibrillogenesis (Koch <i>et al.</i> , 2003).
	Subject:	86004650	GGACCATACCTGC	86004638				
3	gRNA1:	1	GGACCATACCTGC	13	65	No	<i>UCHL5</i>	De-ubiquitinating enzyme (Yao <i>et al.</i> , 2008).
	Subject:	193049737	GGACCATACCTGC	193049749				
4	gRNA1:	1	GGACCATACCTGC	13	65	No	<i>B3GALT2*</i>	Transfers galactose from UDP-galactose to substrates with β GlcNAc residue (Amado <i>et al.</i> , 1998).
	Subject:	194271334	GGACCATACCTGC	194271322				
5	gRNA1:	8	ACCTGCGGAATCG	20	65	No	<i>PGBD5</i>	Transposase related to piggyBac family of proteins (Henssen <i>et al.</i> , 2015).
	Subject:	230359975	ACCTGCGGAATCG	230359963				

* denotes potential off-target sites that were found in UTR sequences or further upstream/downstream of the associated gene's coding sequence.

Next, off-target effects were determined experimentally by Western blotting, but were limited to off-targets identified for *YBX1*sgRNA1. This was because the A2 clone demonstrated the most confident characterised YB1 knockout phenotype by genotypic and phenotypic assays. Additionally, off-target screening was limited to three potential off-targets: PANK4, SOX13 and SDF4 expressions. Firstly, non-transfected 293T harvested at 72h and 12 months in culture served as baseline references for SOX13, PANK4, SDF4 protein expression. The two different harvest points served to exemplify any change in the aforementioned protein expression over the course of 12 months. Considering equal loading was demonstrated by GAPDH loading control, comparable expression patterns of SOX13, PANK4, SDF4 and YB1 levels were detected between 293T control lanes (Fig. 3.17). Therefore, it was generally interpreted that the YB1 knockout clone exhibited comparable levels of SOX13, PANK4 and SDF4 expression throughout the timecourse from passage 5 up to 12 months in culture (Fig. 3.17). This was especially relative to expression levels exhibited by non-transfected 293T control. Overall, Western blot analysis for off-target effects suggested that the CRISPR/Cas9-targeting of *YBX1* gene using *YBX1*sgRNA1 was specific, at least to the limits of detection of Western blotting and for the off-targets selected.

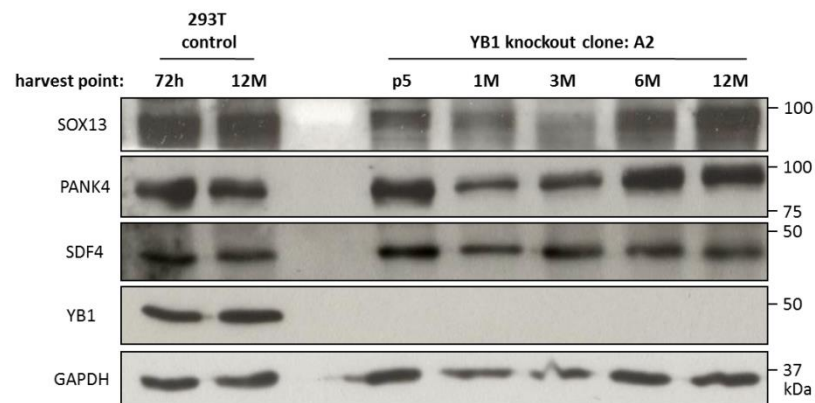


Figure 3.17 Western blot analysis for off-target effects by CRISPR/Cas9 targeting of *YBX1*. Whole cell lysates from non-transfected 293T control and A2 clone cell line were harvested at the indicated harvest points, and samples subjected to Western blotting. Generally, off-target effects were not evident by Western blotting, as SOX13 (top blot), PANK4 (second to top blot) and SDF4 (middle blot) levels were relatively comparable between non-transfected 293T control and A2 clone lanes across the timecourse. This was especially considered after normalised loading by GAPDH loading control (bottom blot). YB1 levels were also assessed (second to bottom blot), and consistent phenotypes were observed for 293T control and YB1 knockout. h, hour; M, month(s); p5, passage 5 post-transfection and single cell cloning of the A2 cell line.

3.2.11 Cell cytotoxicity phenotype was a consequence of YB1 knockout

A key focus of this project was to examine the effect of YB1 knockout on rAAV2 production compared to 293T as producer cells. However, initial attempts to triple transfect clones A2, B2,

or C5 for rAAV2 production using a traditional Calcium Phosphate precipitation method in the presence of chloroquine, was compromised with the observation that YB1 knockout cells showed increased cell death and lysis compared to 293T. In order to establish to which chemical component YB1 knockout cell lines were sensitive to, LDH cytotoxicity assays were performed to quantitatively measure chemical-mediated cell damage in the presence Calcium Phosphate transfection reagents, with and without chloroquine. Here, YB1 knockout clone A2 was used henceforth. The A2 clone was selected to represent the cohort of YB1 knockout cells generated because each clone were observed as equally sensitive to triple transfection for rAAV2 vector production using Calcium Phosphate precipitation with chloroquine, and the lack of CRISPR/Cas9 off target effects for A2 clone was best characterised.

Cytotoxicity in YB1 knockout cells was measured at 24h, 48h, and 72h post-transfection, and presented as %cytotoxicity of max lysis (100%) for the corresponding non-treated control. Non-treated 293T and YB1 knockout control cells functioned as baseline references, and exhibited low and comparable %cytotoxicity – 3-8% throughout the 72h timecourse (Figs. 3.18A-C). In fact, comparable cytotoxicity profiles were displayed between 293T and knockout cells for each treatment condition (no treatment, chloroquine only, Calcium Phosphate mix only, and triple transfection with and without chloroquine) 24h post-treatment (Fig. 3.18A). This indicated that baseline cell cytotoxicities were detected at the earliest time-point measured. However, in the presence of chloroquine, whether standalone or in the context of transfection, YB1 knockout cells demonstrated heightened sensitivity at 48h and 72h post-treatment (Figs. 3.18B and C, respectively). YB1 knockout cells demonstrated average $67.97 \pm 5.17\%$ and $81.46 \pm 8.34\%$ cytotoxicity mediated by chloroquine alone, compared to averages of $42 \pm 6.01\%$ and $48.28 \pm 6.05\%$ cytotoxicity in 293T control, 48h and 72h post-treatment respectively (Figs. 3.18B and C, respectively). Similar average cytotoxicities of $60.66 \pm 8.64\%$ and $79.4 \pm 4.3\%$ for YB1 knockout, and $47 \pm 6.87\%$ and $61.21 \pm 5.57\%$ for 293T control were calculated at 48h and 72h post-transfection with chloroquine, respectively (Figs. 3.18B and C, respectively). The differences in %cytotoxicity between the two cell lines for the given treatment conditions was considered statistically significant at 48h ($n = 3$; $P < 0.001$ and $P < 0.01$ for either conditions, respectively), and 72h ($n = 3$; $P < 0.001$ for both conditions). However, triple transfection without chloroquine displayed reduced cell cytotoxicities that averaged at $11.18 \pm 2.18\%$ and $19.23 \pm 5.65\%$ for 293T control, and $21.54 \pm 0.44\%$ and $31.38 \pm 3.3\%$ for YB1 knockout, at 48h and 72h post-transfection, respectively. The differences in %cytotoxicities between either cell line for each given time-point was also considered

statistically significant ($n = 3$; $P < 0.05$ and $P < 0.01$ for 48h and 72h post-transfection, respectively). It was appreciated that treatment with the Calcium Phosphate mix only (plasmids and chloroquine were omitted) demonstrated comparable cytotoxicity profiles between 293T and YB1 knockout cells only at 48h post-treatment (Fig. 3.18B) with average $8.1 \pm 0.91\%$ and $19.07 \pm 1.5\%$, respectively. 72h post-treatment a statistically significant difference in %cytotoxicity is, however, calculated between 293T control – $9.76 \pm 0.6\%$, and YB1 knockout cells – $22.02 \pm 1.47\%$ (Fig. 3.18C; $n = 3$, $P < 0.05$). The calculated %cytotoxicities were further appreciated with comparable transfection efficiencies displayed for 293T and YB1 knockout cells 72h post-transfection using Calcium Phosphate precipitation supplemented with chloroquine (Fig. 3.18D; $>94\%$ achieved). However, transfections without chloroquine resulted in approximately 88% and 59.2% cells expressing GFP for 293T and YB1 knockout cells, respectively (Fig. 3.18D).

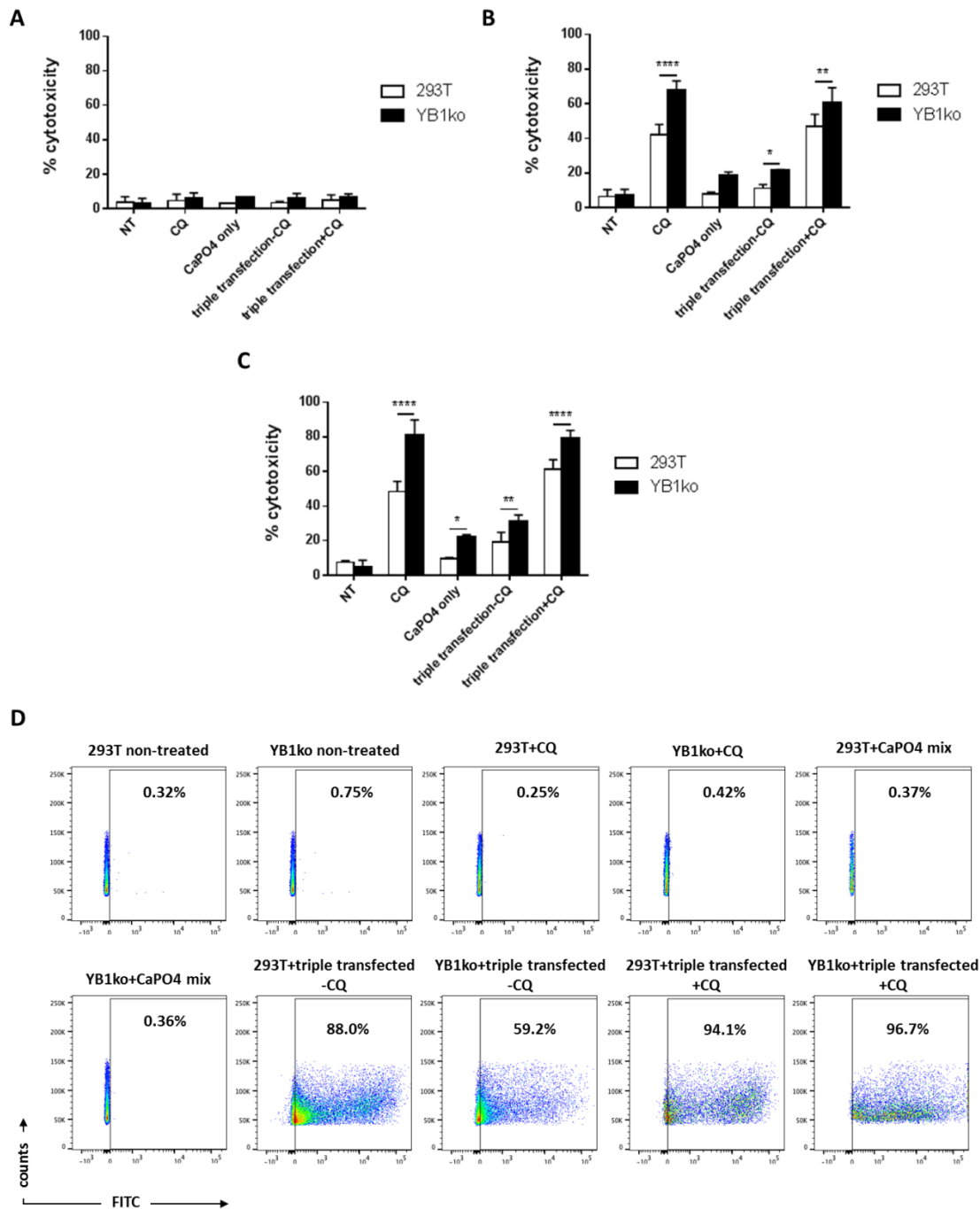


Figure 3.18 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with Calcium Phosphate methodology. Significant LDH activity, represented as %cytotoxicity, was evident in YB1 knockout clones compared to 293T control, after treatment with chloroquine alone or transfection using Calcium Phosphate precipitation method. LDH cytotoxicity profiles were calculated between 293T (solid white bars) and YB1 knockout clone (solid black bars), after treatment with chloroquine (25 μ M), Calcium Phosphate mix, or triple transfection for rAAV2 vector production with and without chloroquine, over a 72h timecourse. LDH activity measured in the cultured media of treated cells at **(A)** 24h, **(B)** 48h, and **(C)** 72h, post-transfection or treatment. Error bars reflect \pm SD from mean, $n = 3$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ or $P < 0.0001$. **(D)** Representative flow cytometry analysis for transfection efficiencies for the pAAV2-hrGFP reporter plasmid 72h post-treatment for 293T and YB1 knockout cell lines, for each condition tested. Negative controls for transfection included 293T and YB1 knockout cells non-treated, and treated with chloroquine or Calcium Phosphate mix, only. Transfection efficiencies are presented within scatter plots as %. CQ, chloroquine; CaPO4 only, Calcium Phosphate mix only (plasmids and chloroquine were omitted); NT, non-treated; YB1ko, YB1 knockout cell line.

Alternative transfection methods are available for the production of rAAV vectors *in vitro*, and include the use of PEI-based reagents for transfection (Huang *et al.*, 2013). 293T and YB1 knockout cells were triple transfected or treated using the alternative PEI_{max} transfection reagent to produce rAAV2 instead, and cytotoxicity profiles examined. We opted to explore whether or not the reported sensitivity to chloroquine and to a lesser extent Calcium Phosphate precipitation method of transfection of YB1 knockout cells extended to other transfection modalities for rAAV vector production.

Therefore, LDH cytotoxicity assays were performed between the 293T and YB1 knockout cell lines with the following conditions: no-treatment, PEI_{max} only, and triple transfection with PEI_{max} for rAAV2 production, over a 72h timecourse. 24h post-treatment, all treatment or transfection conditions demonstrated low and comparable %cytotoxicity at least between 293T and YB1 knockout cells for each given condition, ranging between 2-6% (Fig. 3.19A). Treatment with PEI_{max} only transfection mix (plasmids omitted), showed comparable %cytotoxicities ($n = 3$, $P \geq 0.05$) with average cytotoxicities measured at $6.33 \pm 0.28\%$ and $10.24 \pm 0.17\%$ for 293T, and $9.79 \pm 0.05\%$ and $16.86 \pm 0.34\%$ for YB1 knockout cells, at 48h and 72h post-treatment, respectively (Figs. 3.19B and C, respectively). With respects to the more relevant triple transfection with PEI_{max}, average %cytotoxicities became elevated from $19.31 \pm 7.22\%$ and $21.28 \pm 2.94\%$ for 293T and YB1 knockout cells (respectively) at 48h post-transfection (Fig. 3.19B), to $40.18 \pm 8.46\%$ and $45.48 \pm 5.25\%$ cytotoxicity for 293T and YB1 knockout cells (respectively) 72h post-transfection (Fig. 3.19C). The differences in %cytotoxicities as a result of triple transfection using PEI_{max} between 293T control and YB1 knockout was considered statistically insignificant for either time-points ($n = 3$, $P \geq 0.05$). The calculated %cytotoxicities were further appreciated given that similar transfection efficiencies of 98.4% and 95.5% was exhibited by triple transfected 293T and YB1 knockout cells (respectively) 72h post-transfection (Fig. 3.19D).

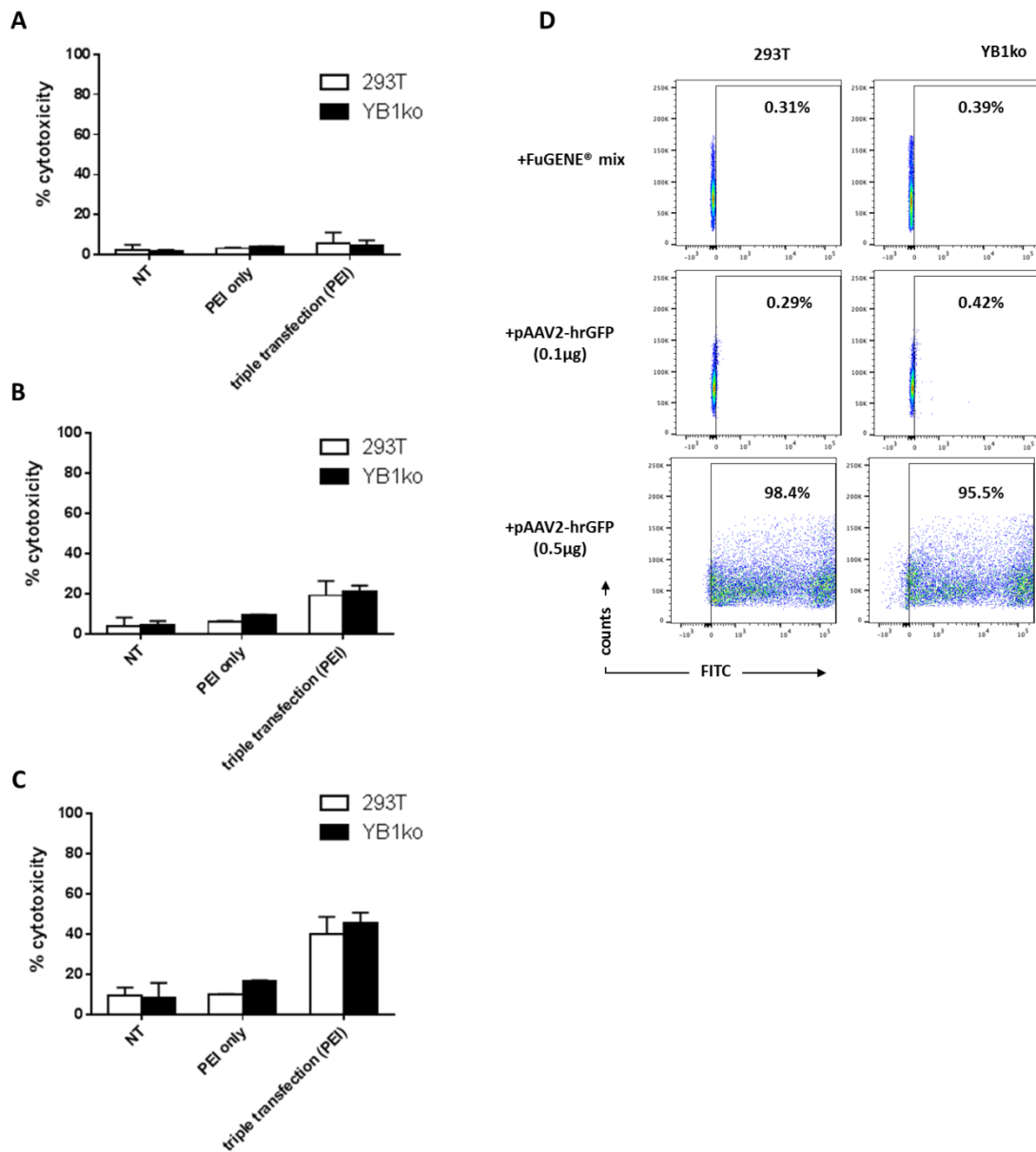


Figure 3.19 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with PEI_{max} reagent. Comparable LDH activity, represented as %cytotoxicity, was evident in YB1 knockout clones compared to 293T control, after treatment with PEI_{max} alone or for transfection for rAAV2 vector production. Results presented indicated that YB1 knockout cells were not significantly susceptible to cell damage or cell death compared to 293T control after treatment or transfection using PEI_{max}, throughout the 72h timecourse. LDH activity was measured in the cultured media between 293T control and YB1 knockout cell lines at **(A)** 24h, **(B)** 48h, and **(C)** 72h, post-transfection or treatment. Error bars reflect \pm SD from mean, n = 3. **D)** Representative flow cytometry analysis for transfection efficiencies for the pAAV2-hrGFP reporter plasmid at 72h post-transfection or -treatment for 293T and YB1 knockout cell lines. Non-treated and PEI_{max} only treated cells functioned as negative controls for flow cytometry analysis. Transfection efficiencies are presented within scatter plots as %. NT, non-treated; YB1ko, YB1 knockout cell line.

Further LDH cytotoxicity assays also were performed using FuGENE® HD and Lipofectamine® transfection reagents over a 72h timecourse (Figs. 3.20 and 3.21, respectively). Cytotoxicity profiles were measured in context to the following conditions: Lipofectamine® or FuGENE®

reagent only (serving as baseline reference control), triple transfection for rAAV2 production, a titration of pAAV2-hrGFP reporter plasmid, and lastly pAAV2-RC plus pHelper plasmids transfected per well, to examine the effect of input plasmid transfection on YB1 knockout cells using alternative transfection modalities. In general, near comparable or no statistically significant difference in cytotoxicity profiles were measured between 293T and YB1 knockout cells for all testing conditions, throughout the timecourse, for either FuGENE® HD or Lipofectamine® transfection modalities (Figs. 3.20A-C and 3.21A-C, respectively; $n = 3$ each, $P > 0.05$). Although, an exception to this was a statistically significant difference ($n = 3$, $P < 0.05$) between 293T and YB1 knockout cells triple transfected with FuGENE® HD reagent (Fig. 3.20C). YB1 knockout cells exhibited an elevated cytotoxicity profile averaging $12.93 \pm 2.63\%$ compared to 293T ($7.34 \pm 1.06\%$), 72h post-transfection.

Generally, Lipofectamine®-based treatments exhibited slightly higher %cytotoxicity compared to FuGENE® HD-treatments, 72h post-transfection (ranging approximately between 20-30% cytotoxicity compared to approximately 10-18% by FuGENE® HD-mediated transfections). Considering the varied range of %cytotoxicity for either cell, especially with regards to Lipofectamine®-based treatment conditions, no real statistical difference in %cytotoxicity was observed between 293T and YB1 knockout cell lines. The general consensus was that %cytotoxicity was comparable between 293T and YB1 knockout cells for most conditions (even at higher reporter plasmid concentrations of pAAV2-hrGFP), because no statistical significant difference between %cytotoxicity was identified at any time-point and for any treatment condition, between 293T and YB1 knockout cells for either Lipofectamine® or FuGENE® HD-mediated treatments. The cytotoxicity profiles were further appreciated by the comparable transfection efficiencies achieved for pAAV2-hrGFP between 293T and YB1 knockout for any given pAAV2-hrGFP transfection condition ($>94\%$), 72h post-transfection. This was recorded using either FuGENE® HD reagents (Fig. 3.20D) or Lipofectamine® (Fig. 3.21D); however, only 85.2% of cells were recorded as GFP-positive after 0.1 μ g input of pAAV2-hrGFP for 293T cells compared to 98.6% for YB1 knockout, 72h post-transfection using FuGENE® HD.

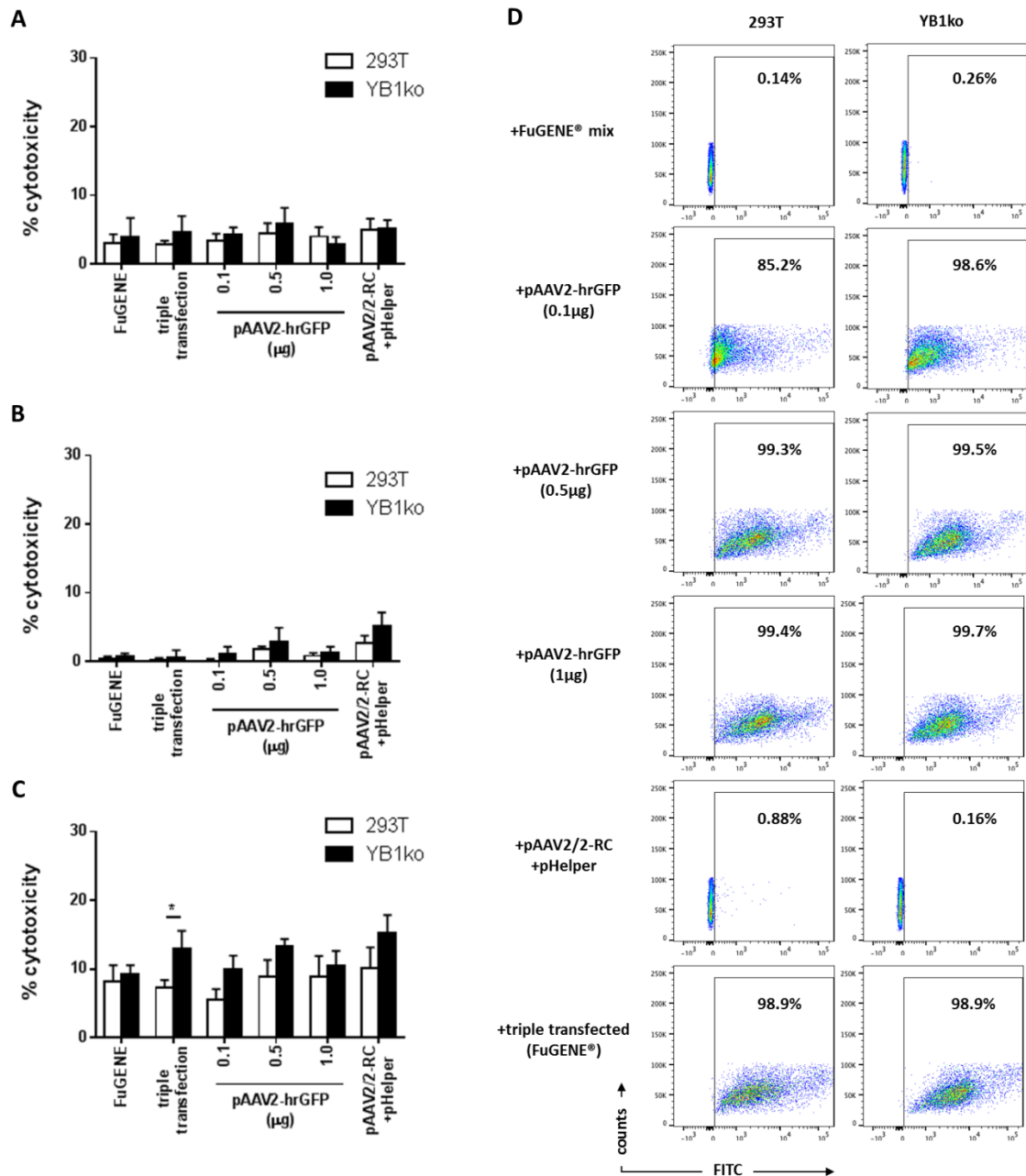


Figure 3.20 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with FuGENE[®] HD reagent. Comparable LDH activity, represented as %cytotoxicity, was evident between YB1 knockout clones and 293T control, after treatment with FuGENE[®] HD alone or transfection with pAAV2-hrGFP, pAAV2/2-RC+pHelper, or triple transfection. Generally, results presented here indicated that YB1 knockout cells were not susceptible to cell damage or cell death compared to 293T control after treatment or transfection using FuGENE[®] HD, throughout the 72h timecourse. LDH activity was measured in the cultured media between 293T control and YB1 knockout cell lines at **(A)** 24h, **(B)** 48h, and **(C)** 72h, post-transfection or treatment. Error bars reflect \pm SD from mean, $n = 3$; *, $P < 0.05$. **D)** Representative flow cytometry analysis for transfection efficiencies for the pAAV2-hrGFP reporter plasmid at 72h post-transfection or -treatment for 293T and YB1 knockout cell lines. Negative controls for flow cytometry included FuGENE[®] HD only treated and pAAV2-RC+pHelper transfections. Transfection efficiencies are presented within scatter plots as %. YB1ko, YB1 knockout cell line.

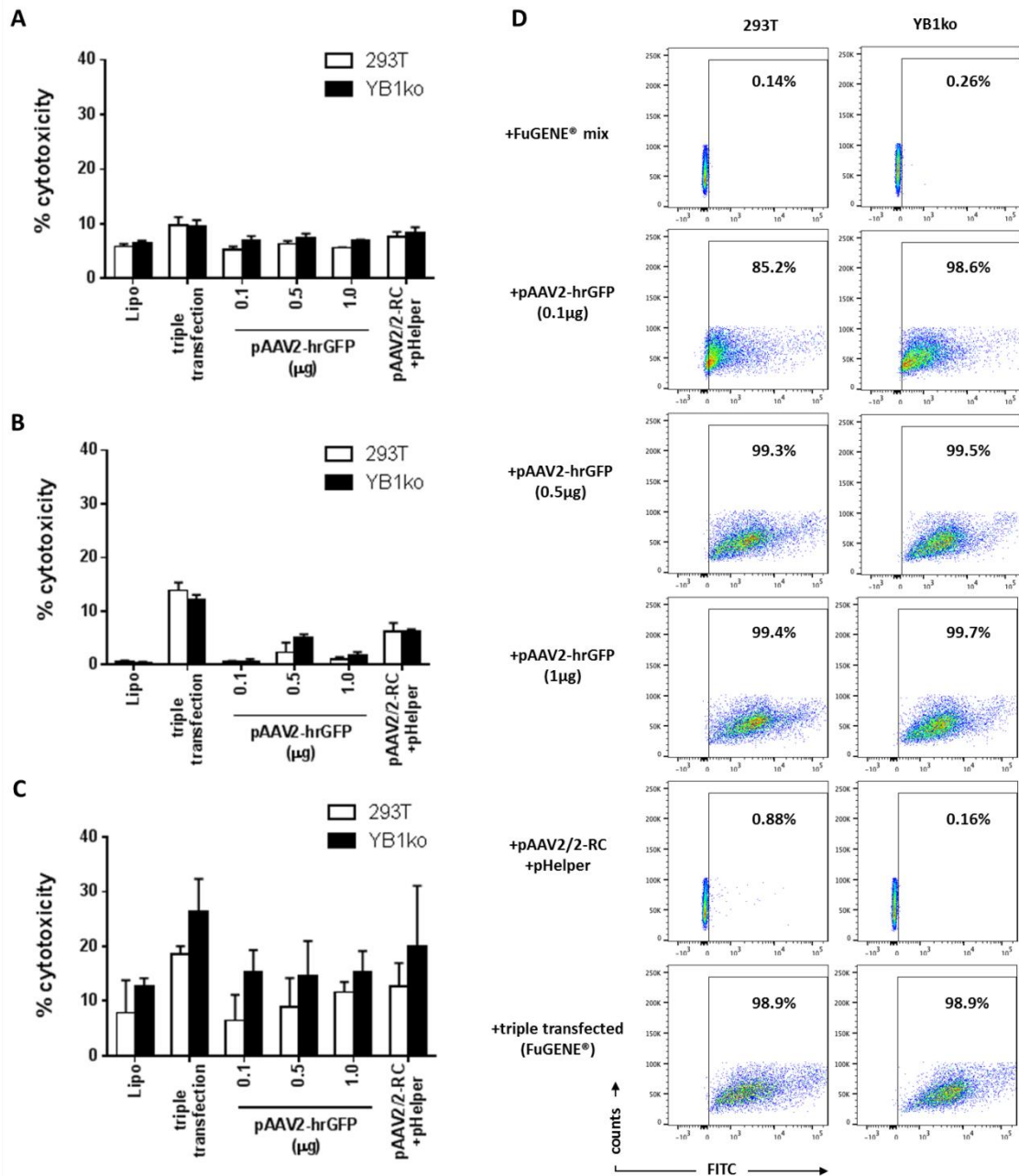


Figure 3.21 LDH cytotoxicity measurements from 293T and YB1 knockout cell lines after transfection with Lipofectamine® 2000 reagent. Comparable LDH activity, represented as %cytotoxicity, was evident between YB1 knockout clones and 293T control, after treatment with Lipofectamine® 2000 alone or transfection with pAAV2-hrGFP, pAAV2/2-RC+pHelper, or triple transfection. Generally, results presented here indicated that YB1 knockout cells were not significantly susceptible to cell damage or cell death compared to 293T control after treatment or transfection using Lipofectamine® 2000, throughout the 72h timecourse. LDH activity was measured in the cultured media between 293T control and YB1 knockout cell lines at **(A)** 24h, **(B)** 48h, and **(C)** 72h, post-transfection or treatment. Error bars reflect \pm SD from mean, n = 3. **D)** Representative flow cytometry analysis for transfection efficiencies for the pAAV2-hrGFP reporter plasmid 72h post-transfection. Negative controls for flow cytometry included Lipofectamine® 2000 only treated and pAAV2-RC+pHelper transfections. Transfection efficiencies are presented within scatter plots as %. Lipo, Lipofectamine® 2000 only; YB1ko, YB1 knockout cell line.

3.2.12 YB1 knockout does not confer an enhanced rAAV2 vector producer cell line

A rAAV2 production system was identified that demonstrated comparable cytotoxicity post-triple transfection between 293T and YB1 knockout cell line – namely PEI_{max}-mediated transfections. Therefore, genome titres of rAAV2 vectors harvested from control 293T and YB1 knockout cell lines were quantified by qPCR. However, the YB1 knockout phenotype did not demonstrate a significant effect on rAAV2 vector genome titres when compared to YB1⁺ 293T control (Fig. 3.22). An equivalent relative vector genome titre difference was measured for YB1 knockout cell line – an average of 0.999 ± 0.405 relative to 293T was calculated (Fig. 3.22A; $n = 4$ batches each cell line quantified, $P \geq 0.05$). Individual physical genome titres per batch were also plotted, and batch-to-batch variation was exemplified when examining individual rAAV2 vector genome titres derived from 293T control and YB1 knockout cell line (Fig. 3.22B). As much as 4.23×10^8 vector genomes/mL was quantified for 293T compared to 5.31×10^8 vector genomes/mL for YB1 knockout (batch 1), corresponding to 1.3X more. Or as low as 2.62×10^8 vector genomes/mL (293T) compared to 1.05×10^8 vector genomes/mL from YB1 knockout cells (batch 2), corresponding to 0.4X less rAAV2 measured from YB1 knockout cells.

The comparable vector genome titres produced from control 293T and YB1 knockout cells was substantiated by comparable transfection efficiencies (>83% for both cell lines), as demonstrated by flow cytometry analysis (Fig. 3.22C). Ultimately, the data presented suggests that CRISPR/Cas9 genome editing for YB1 knockout in 293T cells did not confer an enhanced rAAV2 vector producer cell line in accordance with previously reported YB1 knockdown studies (45-fold enhancement)(Satkunanathan *et al.*, 2014).

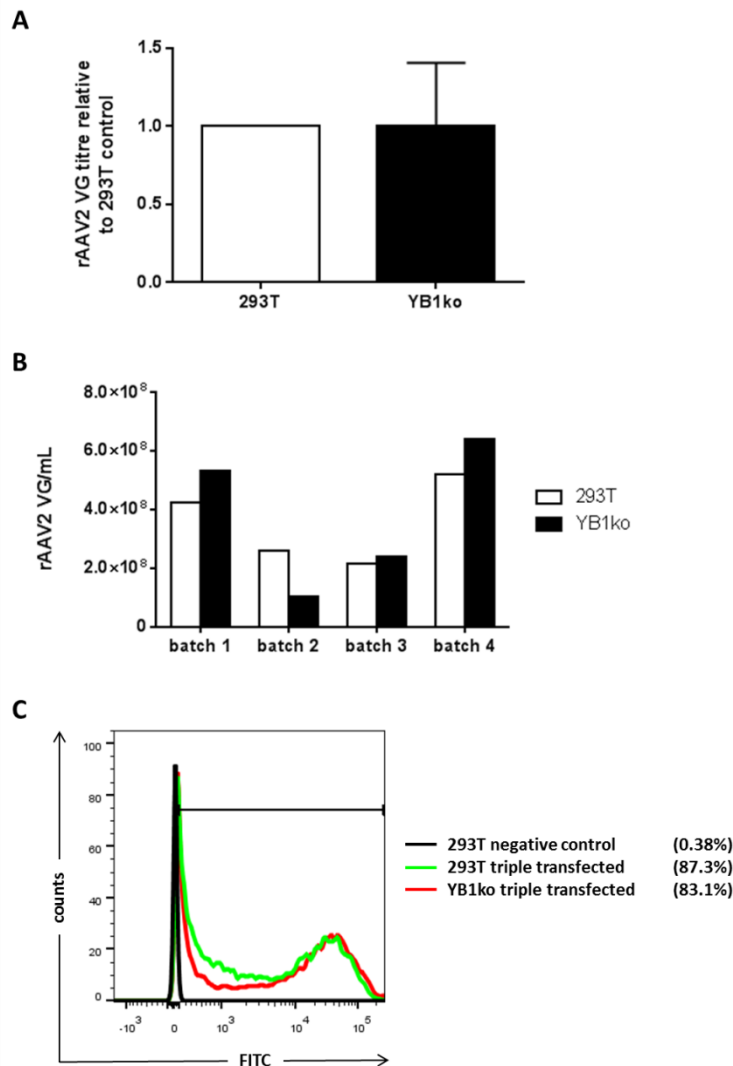


Figure 3.22 Recombinant AAV2 vector genome titres compared between 293T and YB1 knockout producer cell lines by qPCR. Recombinant AAV2 titres between 293T and YB1 knockout cell lines were calculated using qPCR, and titres calculated from YB1 knockout cell line was **(A)** made relative to 293T as the baseline control producer cell line. An equivalent titre difference relative to 293T control producer cell line was calculated. Error bars reflect \pm SD from mean, $n = 4$. **B)** Individual rAAV2 vector genome titres/mL for each batch were plotted and compared between 293T and YB1ko as producer cell lines. **C)** Representative flow cytometry analysis for transfection efficiencies for the pAAV2-hrGFP reporter plasmid at 72h post-transfection of 293T and YB1 knockout cell lines for rAAV2 vector production. Transfection efficiencies are presented in parentheses. VG, vector genome; YB1ko, YB1 knockout cell line.

3.3.0 Chapter Summary

We show in the present study the effective utilisation of CRISPR/Cas9 genome editing tool to generate precise and targeted knockouts of an endogenous gene of choice in 293T cells. Although, given the limitation of 293T's pseudotriploid nature, we demonstrated the benefit of generating single cell clones of pCRISPR-transfected cells to isolate homozygous mutants of the *YBX1* gene. Furthermore, we also show that epitope disruptions can result from genome

editing, and false-positives can be differentiated using a range of target-specific antibodies. However, a main benefit of the present study is the extensive characterisation of CRISPR/Cas9 genome edited cells based on sequencing data, wherein, HRM curve analysis was shown to be a sensitive tool to differentiate wildtype *YBX1* gene sequences from mutated in clones exhibiting a heterozygous genotype, and predictive analysis of alternative splicing motifs and alternative splice site activation was potentially identified. The latter is not a common method of practice utilised in research based on CRISPR technology, currently. However, despite establishing YB1 knockout cell lines using the CRISPR/Cas9 genome editing system and characterising and defining these cell lines at the molecular level as confident YB1 knockouts, the adverse effect of chloroquine on this cell line was unprecedented. This was mainly because chloroquine is commonly used to supplement the Calcium Phosphate precipitation method for transfection (Luthman and Magnusson, 1983; Hasan *et al.*, 1991; Felgner *et al.*, 1994; Kariko *et al.*, 1998). The addition of chloroquine improves transfection efficiency, which is an important reaction to control for optimal viral vector titres. Thus cell cytotoxicity mediated by chloroquine is well tolerable by 293T cells to enable adequate transfection. In addition, no such adverse effect was reported in YB1 knockdown cell lines (Satkunanathan *et al.*, 2014). Therefore, the sensitivity of YB1 knockout cells towards chloroquine treatment was surprising, and prevented the use of this reagent to supplement Calcium Phosphate-based triple-transfection for rAAV vector production. In fact, omission of chloroquine from Calcium Phosphate transfection methodology still presented with a significant difference in cytotoxicity profiles between the two cell lines, even though the overall %cytotoxicity was considered low. In order to confidently assess the impact of YB1 knockout on rAAV vector titres a completely controlled production system, up to and including the transfection dynamics and cytotoxicity profiles between cell lines, was paramount. A difference in cytotoxicity profiles would interfere with data interpretation when comparing the effect of YB1 knockout on AAV processing. However, examination of alternative transfection reagents revealed that PEI-mediated transfection, and to a lesser extent Calcium Phosphate (without the supplementation of chloroquine), for rAAV vector production remains a suitable alternative. Despite this, overall rAAV2 vector genome titres was not enhanced in the current study using YB1 knockout cell lines compared against un-edited 293T cells as baseline control, which was in contrast to an analogous study that demonstrated that downregulation of YB1 in 293T cells correlated with an enhanced rAAV2 and rAAV8 vector genome titres (Satkunanathan *et al.*, 2014). The possible reasons for which are discussed in detail in Chapter 7 (see sections 7.7,

7.9, and 7.11 for further discussions). In the next chapter, we further investigate the impact and the role of YB1 and YB1 knockout has on rAAV2 vector production.

Chapter 4: The Role of YB1 on Recombinant AAV Vector Production

4.1.0 Introduction

A novel approach to improve rAAV vector systems is the manipulation and regulation of endogenous, cell-intrinsic protein expression to influence rAAV vector production and output. Recently, host cell factors, including Y-Box protein (YB)1 protein, have been found associated with rAAV particles; and targeted knockdown of YB1 protein expression in 293T cells correlated with transiently enhanced rAAV vector titres and Rep protein expression (Satkunanathan *et al.*, 2014). In Chapter 3 we employed CRISPR/Cas9 genome editing for targeted knockout of YB1 expression in 293T cells and explored the feasibility of this approach on rAAV vector production. We now expand on the YB1 knockout cell line, and the aim of the current study was to further examine the role of YB1 in context of rAAV vector processing. Having previously established YB1 knockout cell lines using CRISPR/Cas9 genome editing, we aimed to elucidate the capacity at which YB1 expression potentially rescues the cytotoxic phenotype associated with selected transfection modalities. Furthermore we were also interested in which functional domain of YB1 contributes to rAAV vector production dynamics, including rAAV2 vector genome titres, AAV2 protein localisation and expression. Overall, we demonstrated that the chloroquine-induced cell cytotoxicity could be rescued in YB1 knockout cell lines by the stable expression of full length and select YB1 mutants. Additionally, an insignificant effect on rAAV2 vector processing was evident by the YB1 knockout cells and the stable expression of a range of YB1 truncation mutant phenotypes. Despite this, putative YB1 DNA binding motifs were identified in AAV and Adenovirus sequences, to which specific binding to the *ITR*, *rep* and *E2a* DNA sequences was revealed, potentiating in YB1 having an effect of these elements *in vitro*.

4.2.0 Results

4.2.1 Exogenous YB1 expression by plasmid transfection of 293T and YB1 knockout cell lines

We previously demonstrated that YB1 knockout cells were susceptible to cell cytotoxicity after transfection and/or treatment with chloroquine reagent (Chapter 3). We next examined the

feasibility in reintroducing exogenous YB1 using pDESTmycYBX1 transfection to both 293T and YB1 knockout cells. Therefore, cells were transfected with various amounts of pDESTmycYBX1 and cultured without antibiotic selection. Cell cytotoxicity or any adverse effect towards pDESTmycYBX1 transfection was not observed, and so Western blotting was performed to determine the expression levels of endogenous and exogenous YB1 upto 30 days in culture, post-transfection (Fig. 4.1).

Given equal loading of protein lysates was well represented by GAPDH across the entire timecourse, overall, exogenous YB1 (myc-tagged) was markedly expressed in 293T and YB1 knockout cells 72h post-transfection (Figs. 4.1A and B, respectively). This was particularly true for input quantities of 1 μ g, 2 μ g, and 5 μ g of pDESTmycYBX1 plasmid. Exogenous YB1-myc expression was considered easily distinguished from endogenous YB1 – the former resolved to approximately 52kDa and endogenous YB1 to approximately 50kDa (Figs. 4.1A and B, compare YB1-myc annotation to YB1 annotation). However, the expression of exogenous YB1-myc was eventually lost in both 293T and YB1 knockout cells, even when transfected with the highest quantities of pDESTmycYBX1 as early as 10 days post-transfection (Figs. 4.1C and D, respectively), and at 20 days (Figs. 4.1E and F, respectively), and finally 30 days (Figs. 4.1G and H, respectively) post-transfection. Non-transfected 293T and YB1 knockout cells functioned as negative controls for pDESTmycYBX1 transfection, and throughout the timecourse did not exhibit exogenous YB1-myc expression, especially at the 72h harvest point (Figs. 4.1A-H). On the other hand, pDESTmycYBX1-transfected YB1 knockout cells recovered their knockout status after the loss of YB1-myc expression after 72h post-transfection (Figs. 4.1D, F, and H). In conclusion, exogenously introduced YB1 expression in 293T and YB1 knockout cell lines showed no detrimental effects, but stable transformation of either cell line was not achieved using pDESTmycYBX1 plasmid transfection, especially without antibiotic selection.

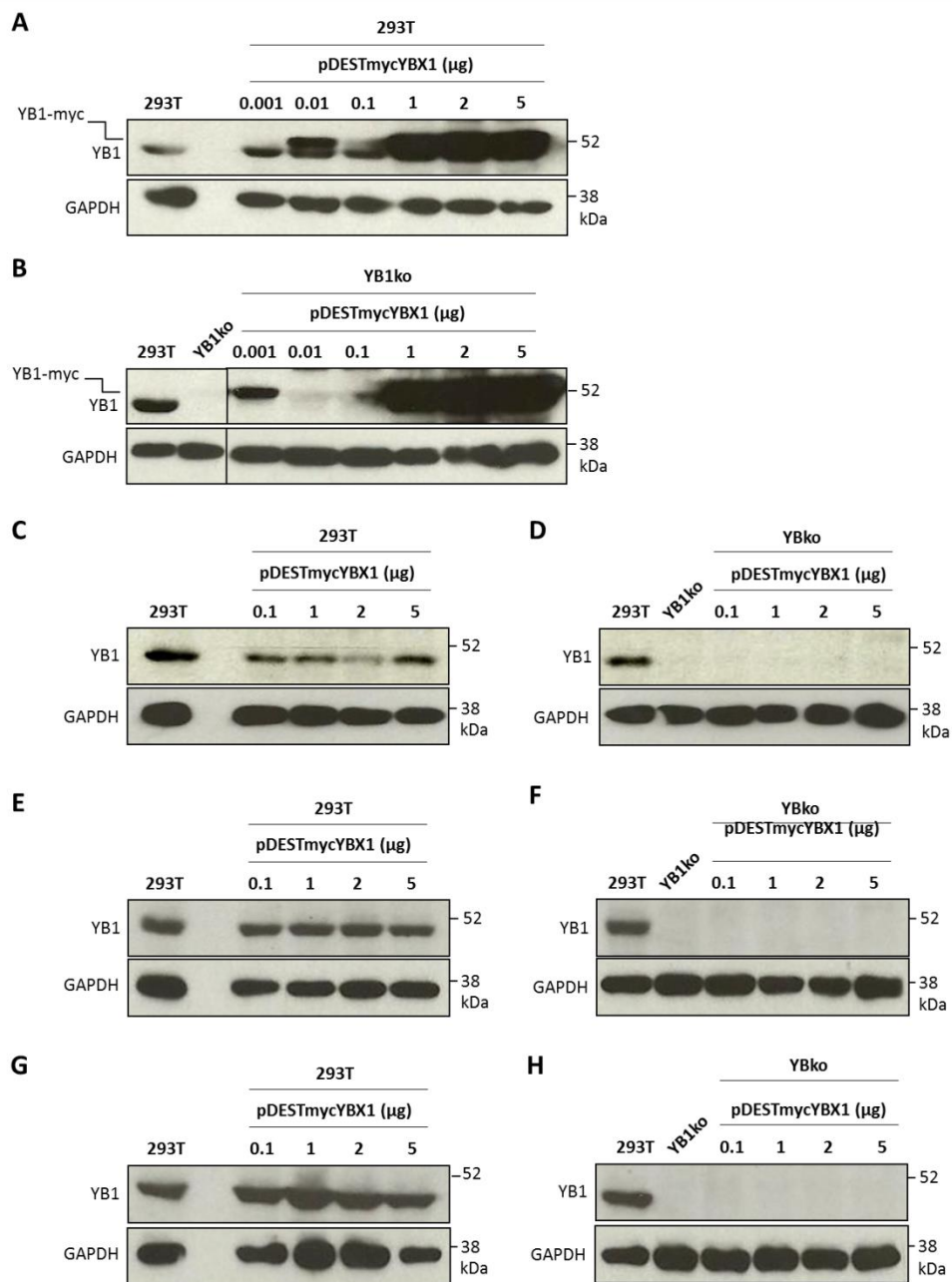


Figure 4.1 Western blot analysis of exogenous YB1 expression in 293T and YB1 knockout cell lines after pDESTmycYBX1 transfection. Reintroducing and overexpressing YB1 was investigated by transfecting 293T and YB1 knockout (A2) cells at different quantities (μg) of pDESTmycYBX1. Western blot analysis demonstrated overexpression and reintroduction of YB1 expression (top blots; see YB1-myc annotated signal compared to endogenous YB1), 72h post-transfection between **(A)** 293T and **(B)** YB1 knockout cells. However, the expression of exogenous YB1 was considered transient in both 293T and YB1 knockout cells, as by **(C and D, respectively)** 10 days, **(E and F, respectively)** 20 days, and **(G and H respectively)** 30 days post-transfection, exogenous YB1-myc expression was lost. Baseline YB1 expression was restored for **(C, E, G)** 293T. Similarly, **(D, F, H)** YB1 knockout cells recovered the YB1 knockout phenotype. Non-transfected 293T and YB1 knockout cells served as negative controls for transfection. GAPDH expression (bottom panels) was used to determine equal loading. YB1ko, YB1 knockout cell line.

4.2.2 Infection of YB1 knockout cells with rAAV or lentiviral vectors exhibit comparable cytotoxicity profiles to 293T control

We next explored if the YB1 knockout phenotype exhibited adverse effects to viral vector (rAAV2 and lentiviral vectors) infections at various multiplicities of infection (MOI) in comparison to 293T as control, using LDH cytotoxicity assays. As positive controls for LDH cytotoxicity, 293T and YB1 knockout cells were also triple transfected for rAAV2 vector production using Calcium Phosphate precipitation method supplemented with chloroquine in parallel. As a result, comparable %cytotoxicity profiles were measured in response to rAAV2GFP vector infections at each tested MOI (10, 100, and 1000), throughout the entirety of the 72h timecourse (Figs. 4.2A-C; $n = 3$, $P \geq 0.05$). Infections with rAAV2GFP did not exceed the %cytotoxicity exhibited by no-treatment controls, which averaged $4.34 \pm 1.87\%$ and $4.16 \pm 1.76\%$ for 293T and YB1 knockout cells, respectively, 72h post-transfection (Fig. 4.2C). On the other hand, triple transfection using Calcium Phosphate precipitation method with chloroquine resulted in significant cytotoxicity averaging at $21.1 \pm 12.31\%$ and $70.57 \pm 13.87\%$ for YB1 knockout cells, compared to $7.71 \pm 0.76\%$ and $36.1 \pm 5.8\%$ for 293T control, 48h and 72h post-transfection, respectively (Figs. 4.2B and C; $n = 3$, $P < 0.01$ and $P < 0.001$, respectively). Additionally, cytotoxicity profiles were considered relatively independent of vector transduction rates, as measured by flow cytometry 72h post-infection (Fig. 4.2D). Transduction efficiency at a given MOI was incomparable between 293T and YB1 knockout cells; for example, using MOI 100, approximately 3.59% and 1.89% of 293T and YB1 knockout cells, respectively, were GFP-positive. The disparity was more pronounced at the higher MOI 1000, where an appreciable 30.3% and 13.2% of 293T and YB1 knockout cells, respectively, were GFP-positive (Fig. 4.2D). The comparable cytotoxicities after rAAV2GFP infection were therefore considered independent of vector transduction. Overall, despite exhibiting equivalent cytotoxicity profiles to 293T control, the YB1 knockout cells interestingly demonstrated a poor transduction capacity for rAAV2 vector compared to 293T.

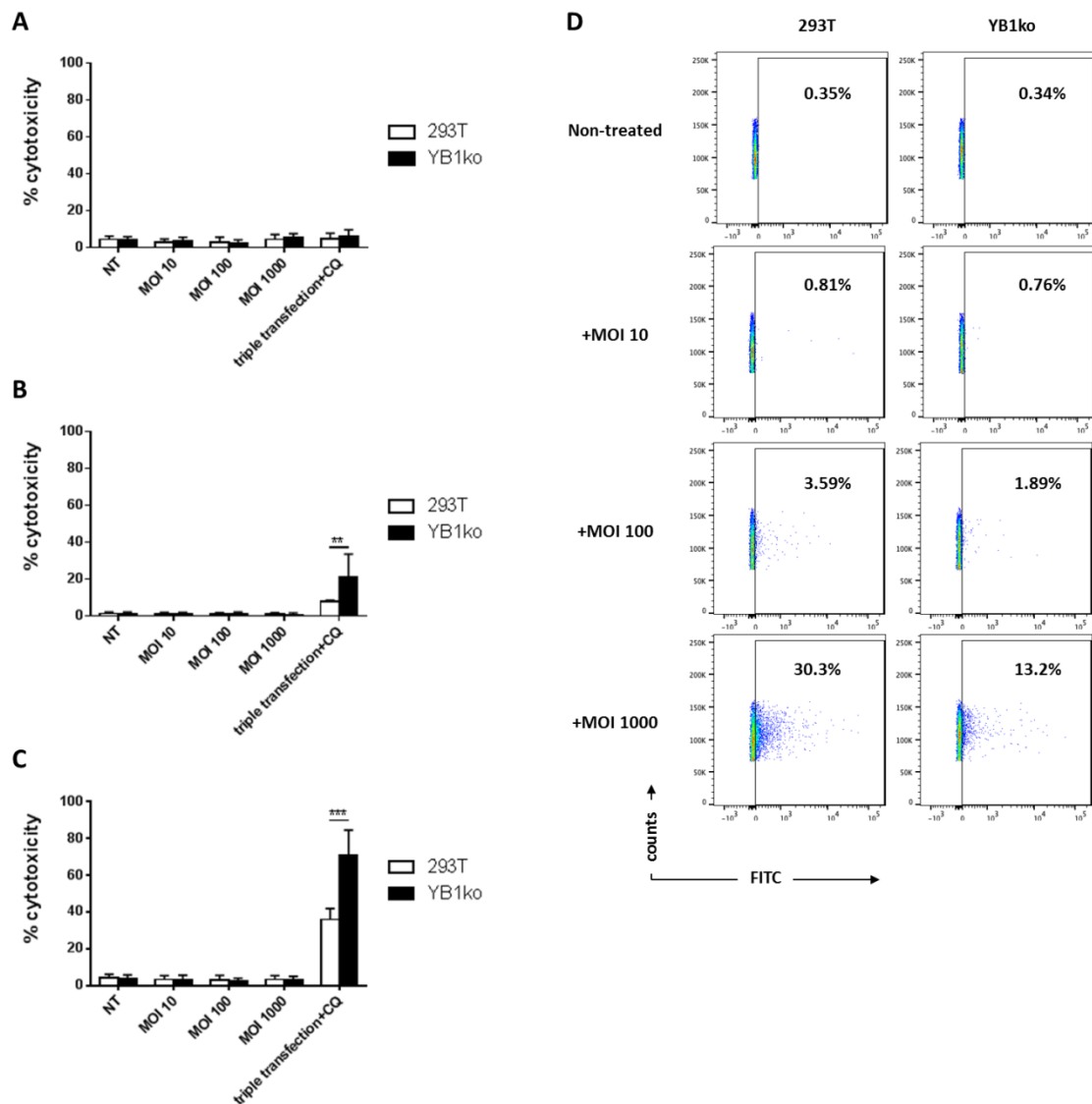


Figure 4.2 LDH cytotoxicity assays for rAAV2GFP transduction of 293T and YB1 knockout cells. 293T and YB1 knockout cells were infected with rAAV2GFP vectors at the indicated MOIs, and cytotoxicity in response to rAAV2GFP transduction was measured for no-treatment and MOIs 10, 100, and 1000, over a 72h timecourse. LDH activity was measured at (A) 24h, (B) 48h, and (C) 72h, post-infection, showing overall that YB1 knockout cells exhibited comparable cytotoxicity profiles to 293T control cells at all MOIs. Triple transfection for rAAV2GFP production using Calcium Phosphate precipitation method, supplemented with chloroquine was used as a positive control for LDH activity, and measured in parallel. Error bars reflect \pm SD from mean, $n = 3$; **, $P < 0.01$; ***, $P < 0.001$. D) Flow cytometry measured the transduction efficiency between 293T and YB1 knockout 72h post-infection with rAAV2GFP. No-treatment control served as negative control for flow cytometry. Transduction efficiencies are presented within scatter plot. CQ, chloroquine; MOI, multiplicity of infection; NT, no-treatment control; YB1ko, YB1 knockout cell line.

Similar was observed and calculated for LDH cytotoxicity assays after infections of 293T and YB1 knockout cells using lentiviral vector (lentiGFP) at MOIs 1, 5 and 30, over a 72h timecourse. Comparable cytotoxicity profiles between 293T and YB1 knockout cell lines were observed for each tested MOI of lentiGFP and control treatments (no-treatment control, opti-MEM™, and opti-MEM™ supplemented with polybrene), across the entirety of the 72h

timecourse (Figs. 4.3A-C; $n = 3$, $P \geq 0.05$). C). In fact, %cytotoxicity after lentiGFP infections or control treatments also did not exceed %cytotoxicity exhibited by non-treatment controls, which averaged $5.14 \pm 1.26\%$ and $6.88 \pm 1.32\%$ for 293T and YB1 knockout cells, respectively, 72h post-transfection (Figs. 4.3C). Thus implying that, likewise with rAAV2GFP infections, no adverse effects were measurably detected in YB1 knockout cell line in response to opti-MEM™, polybrene or lentiviral vectors infections at the MOIs tested. Analyses were further appreciated given the triple transfection using Calcium Phosphate precipitation method with chloroquine resulted in significant cytotoxicity averaging at $40.25 \pm 3.87\%$ and $68.82 \pm 6.73\%$ for YB1 knockout cells, compared to $11.96 \pm 1.6\%$ and $32.17 \pm 2.73\%$ for 293T control, 48h and 72h post-transfection, respectively (Figs. 4.3B and C; $n = 3$, $P < 0.001$ for both time-points). Transduction efficiencies of lentiGFP infection of 293T and YB1 knockout cell lines were also measured using flow cytometry 72h post-infection (Fig. 4.2D). In contrast to the disparity in transduction efficiencies measured after rAAV2GFP infections between 293T and YB1 knockout cells, transduction efficiencies were comparable between the two cell lines for the given MOI 72h post-infection. Overall, an equivalent or similar transduction capacity for lentiGFP vector between 293T and YB1 knockout cells was concluded. Nonetheless, both viral vector systems were considered permissible for transduction of YB1 knockout cells, and contributed to no measurable adverse effect, especially when compared to 293T control cell line.

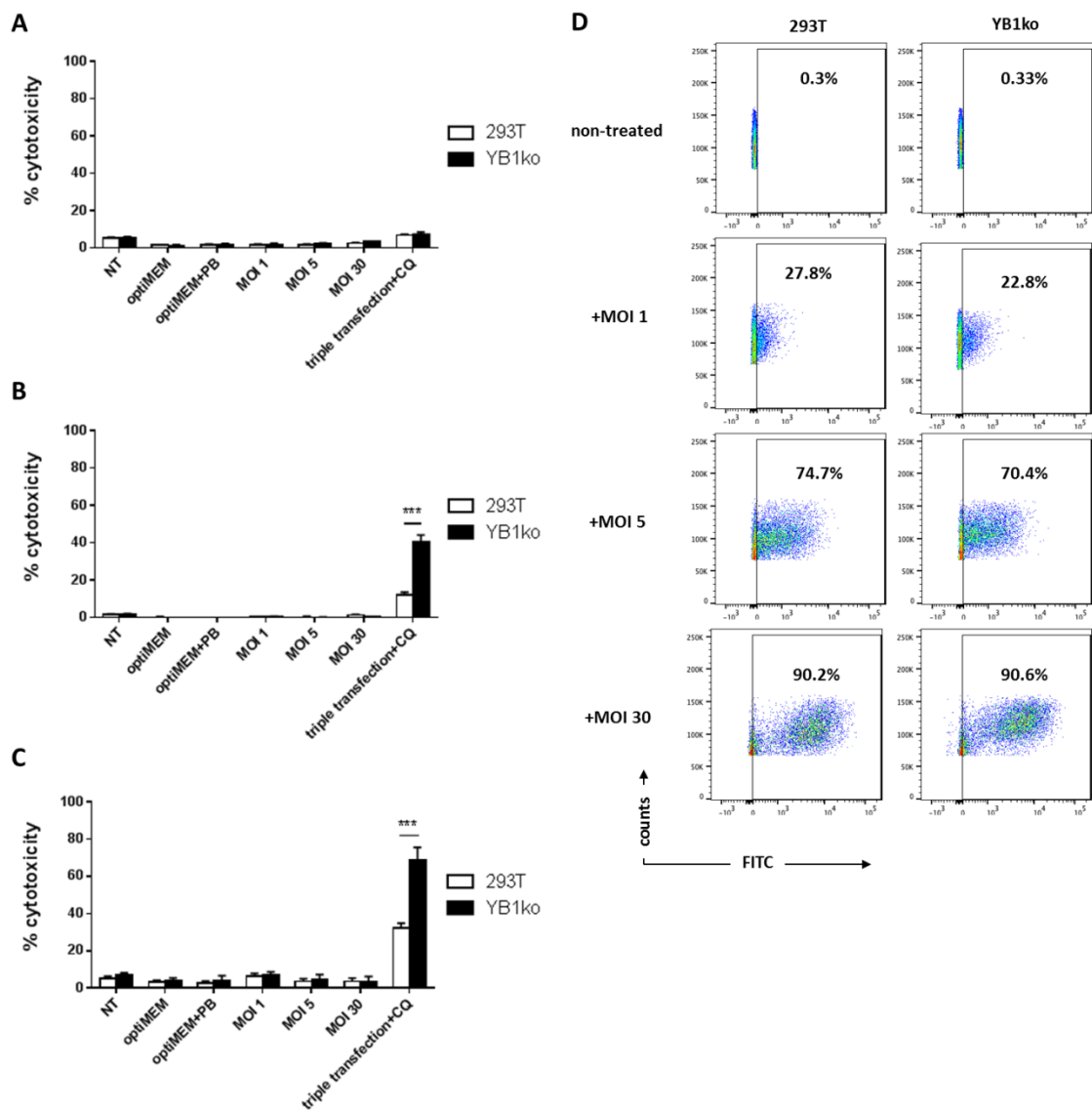


Figure 4.3 LDH cytotoxicity assays for lentiviral vector (lentiGFP) transduction of 293T and YB1 knockout cells. 293T and YB1 knockout cells were infected with lentiGFP vector at the indicated MOIs. Cell cytotoxicity was measured using LDH cytotoxicity assays, for no-treatment, conditioning with optiMEM, optiMEM with polybrene controls, and MOIs 1, 5, and 30 of lentiGFP over a 72h timecourse. LDH activity was measured at **(A)** 24h, **(B)** 48h, and **(C)** 72h post-infection, showing overall that YB1 knockout cells exhibited comparable cytotoxicity profiles to 293T control for all MOIs. Triple transfection for rAAV2GFP production using Calcium Phosphate precipitation method supplemented with chloroquine was used as a positive control for LDH activity. Error bars reflect \pm SD from mean, $n = 3$; *******, $P < 0.001$. **D**) Flow cytometry demonstrated comparable infection efficiencies for each MOI between 293T and YB1 knockout 72h post-infection. No-treatment control functioned as negative control for flow cytometry. Transduction efficiencies are presented in parentheses. CQ, chloroquine; MOI, multiplicity of infection; NT, no treatment control; PB, polybrene; YB1ko, YB1 knockout cell line.

4.2.3 Generating stable cell lines expressing YB1 full length and truncation mutants

In order to dissect the functional domain responsible for the protective phenotype to chloroquine-induced cytotoxicity in 293T control cells, YB1 mutants were reintroduced into the YB1 knockout cell line. The aforementioned LDH cytotoxicity assays indicated that

lentiviral vectors were optimal for gene transfer in YB1 knockout cells. Therefore, full length and YB1 mutants were designed for stable expression (Fig. 4.4A). In short, full length YB1 (YB1FL) functioned to rescue YB1 expression, whereas truncated mutants (YB1 Δ 1- Δ 6) intended to examine separate functional domains of the YB1 protein. Cloning fragments that encoded the intended YB1 protein forms were PCR amplified from pDESTmycYBX1 template, and the approximate sized DNA fragments that were expected after agarose gel electrophoresis were identified (Fig. 4.4B). The approximate DNA fragments corresponded to *YB1FL* (1004bp), *YB1 Δ 1* (799bp), *YB1 Δ 2* (682bp), *YB1 Δ 3* (418bp), *YB1 Δ 4* (532bp), *YB1 Δ 5* (617bp), and *YB1 Δ 6* (853bp). Restriction enzyme digestions of pDUAL.mIL6-puro lentiviral transfer plasmid demonstrated *mIL6* dropout (approximately 655bp; see red arrow), which was distinct from single and uncut controls (Fig. 4.4C; see +BamHI+NotI lane labelled prior and red arrow for *mIL6* dropout, and compare to controls). Furthermore, the pDUAL vector backbone post-gel purification was double restriction digested and demonstrated that a relatively clean pDUAL vector backbone was isolated, with no detectable *mIL6* dropout (Fig. 4.4C; see lane post and compare to lane prior). Unfortunately, two alternative YB1 truncations were not possible to clone in the present study due to cloning difficulties. These included constructs reflecting only the A/P domain or the A/P plus the CTD (i.e. absent of the CSD).

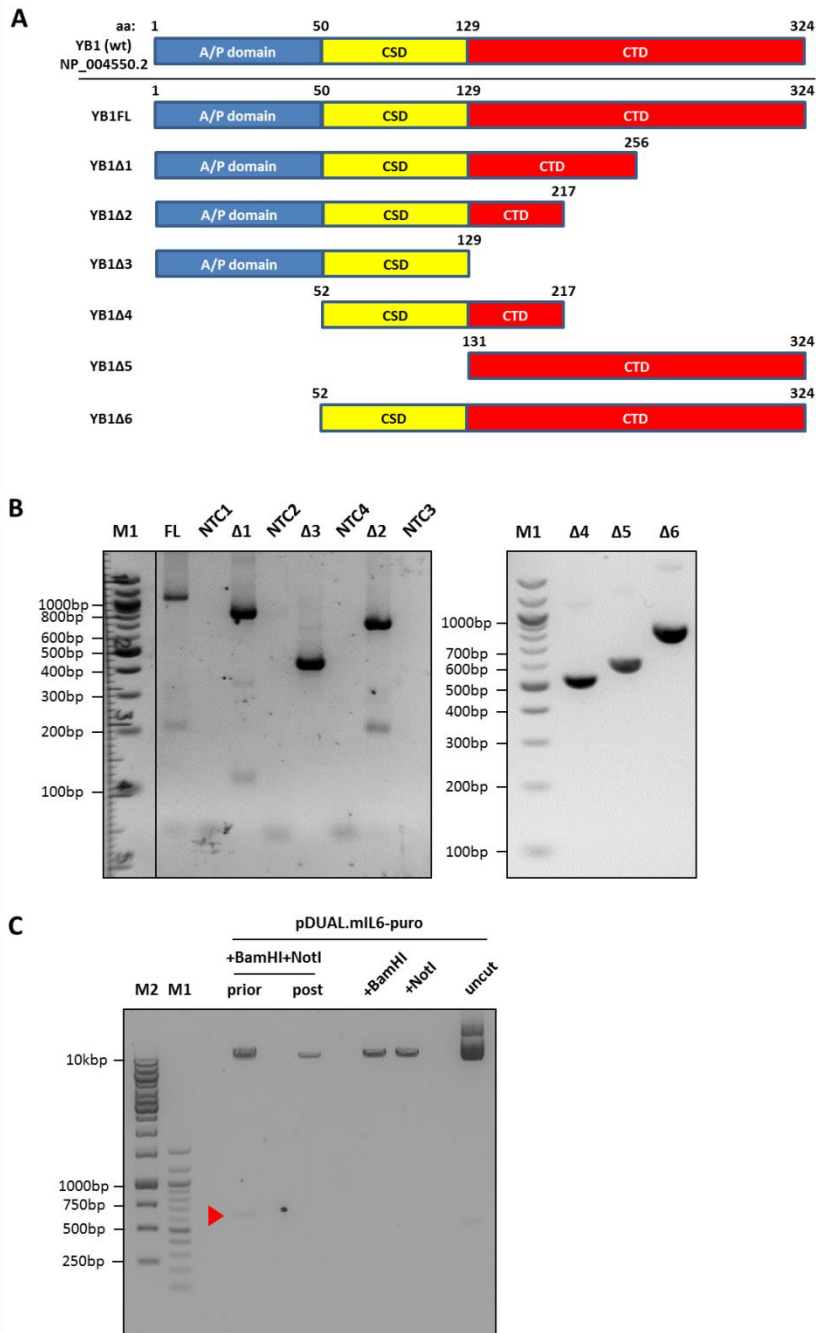


Figure 4.4 Designing full length and mutant YB1 coding sequences for cloning into pDUAL vector backbone for lentiviral vector production. A) Schematic of full length YB1 (YB1FL) and YB1 mutants (YB1Δ1-Δ6) intended to be stably expressed from YB1 knockout cells. PCR amplifications of full length *YBX1* (YB1FL) and mutant YB1-encoding sequences (YB1Δ1-Δ6) were (B) confirmed by 1% TAE agarose gel electrophoresis. Expectant PCR amplicons of approximate sizes were verified (YB1FL, 1004bp; YB1Δ1, 799bp; YB1Δ2, 682bp; YB1Δ3, 418bp; YB1Δ4, 532bp; YB1Δ5, 617bp; YB1Δ6, 853bp). C) The quality of pDUAL.mIL6-puro vector backbone prep was also verified by 1% TAE agarose gel electrophoresis. Dropout of the *mIL6* fragment (approximately 655bp, see red arrow) was distinguished from the pDUAL.mIL6-puro vector backbone prior to gel purification (lane prior), and post-purification (post). Digestion was compared to single and uncut plasmid controls. aa, amino acid; M1, marker lane, 100bp DNA Ladder; M2, GeneRuler 1kb DNA Ladder; NTC, non-template control; wt, wildtype.

The *YB1FL* and *YB1Δ1-Δ6* DNA fragments were cloned into the pDUAL lentiviral vector backbone. Analytical restriction enzyme digestion revealed that the approximate sized DNA fragments corresponded to *YB1FL* (992bp), *YB1Δ1* (788bp), *YB1Δ2* (671bp), *YB1Δ3* (407bp), *YB1Δ4* (521bp), *YB1Δ5* (605bp), and *YB1Δ6* (842bp), for most, if not all, selected transformants per cloning (Fig. 4.5A). Restriction enzyme digestion of the parental pDAUL.mIL6-puro plasmid functioned to differentiate false-positive transformants, and with a *mIL6* 655bp DNA fragment dropout, it was difficult to conclusively differentiate DNA fragments derived from +*YB1Δ2* transformants or +*YB1Δ5* transformants. Nonetheless, cloning was further confirmed by Sanger sequencing and results are depicted in Figs. 4.5-4.7 (Figs. 4.5B and C, *YB1FL* and *YB1Δ1*, respectively; Figs. 4.6A-C, *YB1Δ2-4*, respectively; and Figs. 4.7A and B, *YB1Δ5* and *YB1Δ6*, respectively). Cloning junctions at both 5'- and 3'- ends of the *YB1FL* or *YB1Δ1-Δ6* sequences were correctly presented. Furthermore, cloning fragments corresponded to the *YB1FL*, or *YB1Δ1-Δ6* sequences were all in-frame with their respective ATG start codon (Figs. 4.5-4.7; see top sequencing panels – SFFV.F(oward) derived sequencing, green underlined sequence). Additionally, TAA stop codons were in-frame with their respective *YB1FL* or *YB1Δ1-Δ6* coding sequences (Figs. 4.5-4.7; see bottom sequencing panels – WPRE.R(everse) derived sequencing, green underlined sequence). Although, sequencing of the selected +*YB1Δ3* transformant using the WPRE.R(everse) primer was not determined (Fig. 4.6B; bottom sequencing panel), the expected 407bp DNA fragment dropout of approximate size was successfully identified by analytical restriction enzyme digestion (Fig. 4.5A). These lentiviral transfer plasmids were referred to as pDUAL.YB1FL or pDUAL.YB1Δ1-Δ6, and used to generate lentiviral vectors.

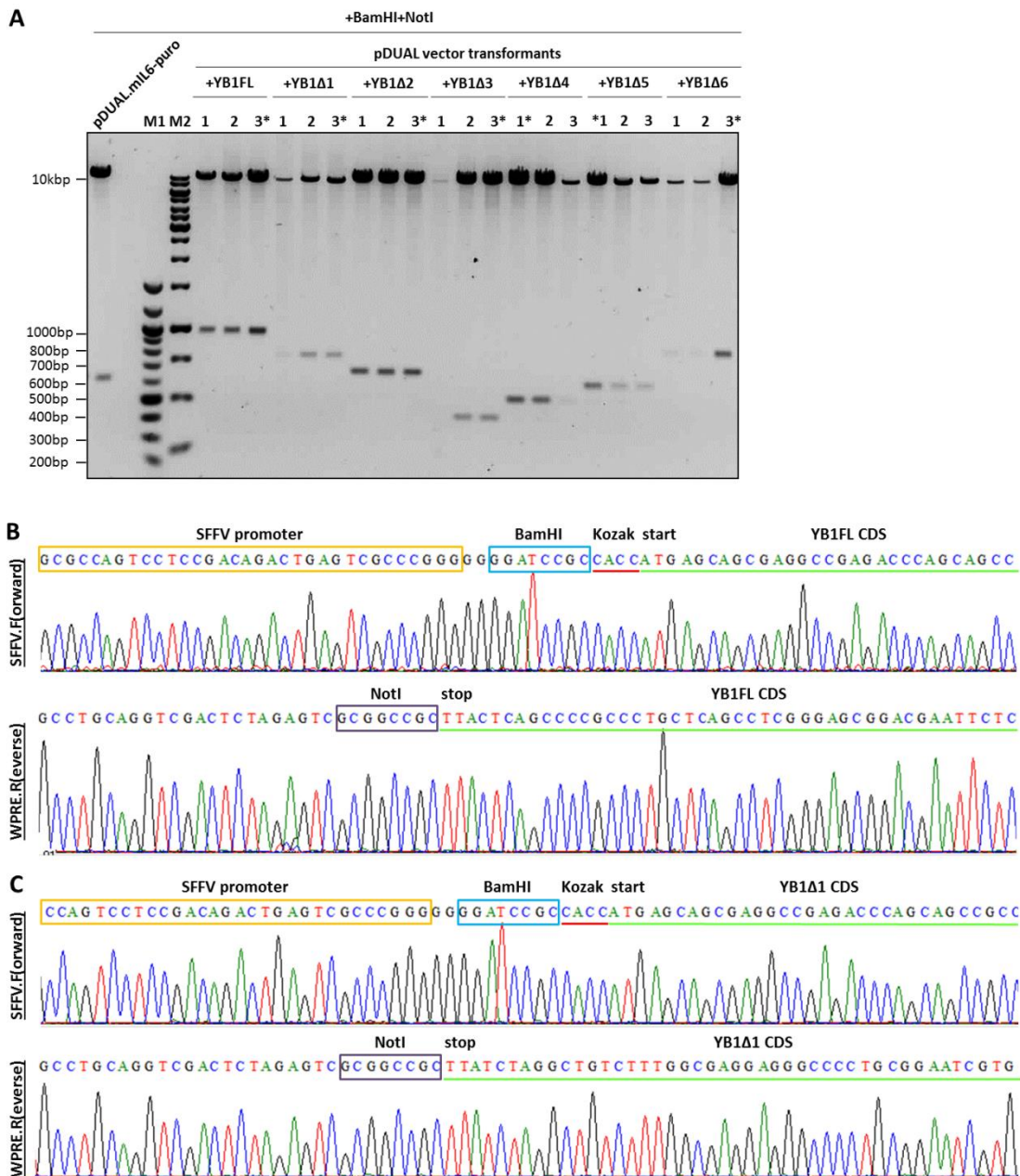


Figure 4.5 Cloning of *YBX1* full length and mutant *YB1* sequences into pDUAL vector backbone for lentiviral vector production. *YB1FL* and *YB1Δ1-Δ6* sequences were cloned into pDUAL vector backbone and **(A)** analytical restriction enzyme digest and 1% TAE agarose gel electrophoresis confirmed positive transformants with approximate sized DNA fragments (*YB1FL*, 1004bp; *YB1Δ1*, 799bp; *YB1Δ2*, 682bp; *YB1Δ3*, 418bp; *YB1Δ4*, 532bp; *YB1Δ5*, 617bp; *YB1Δ6*, 853bp). Double digested pDUAL.mil6-puro was also run in parallel to help differentiate false-positive screens. **B-H)** Sequence and sequence chromatograms demonstrated successful cloning for selected transformants for **(B)** *YB1FL*, and **(C)** *YB1Δ1* into the pDUAL lentiviral transfer vector backbone, using SFFV.F(orward) and WPRE.R(verse) sequencing primers (top and bottom panels, respectively). Coding sequences were considered in-frame with their start codon. SFFV promoter (yellow boxed), BamHI (blue boxed), Kozak sequence (red underlined), coding sequence, green underlined, ATG codon, start, TAA reverse complement codon, stop, and NotI (grey boxed) sequences are annotated. M1, 100bp DNA Ladder; M2, GeneRuler 1kb DNA Ladder.

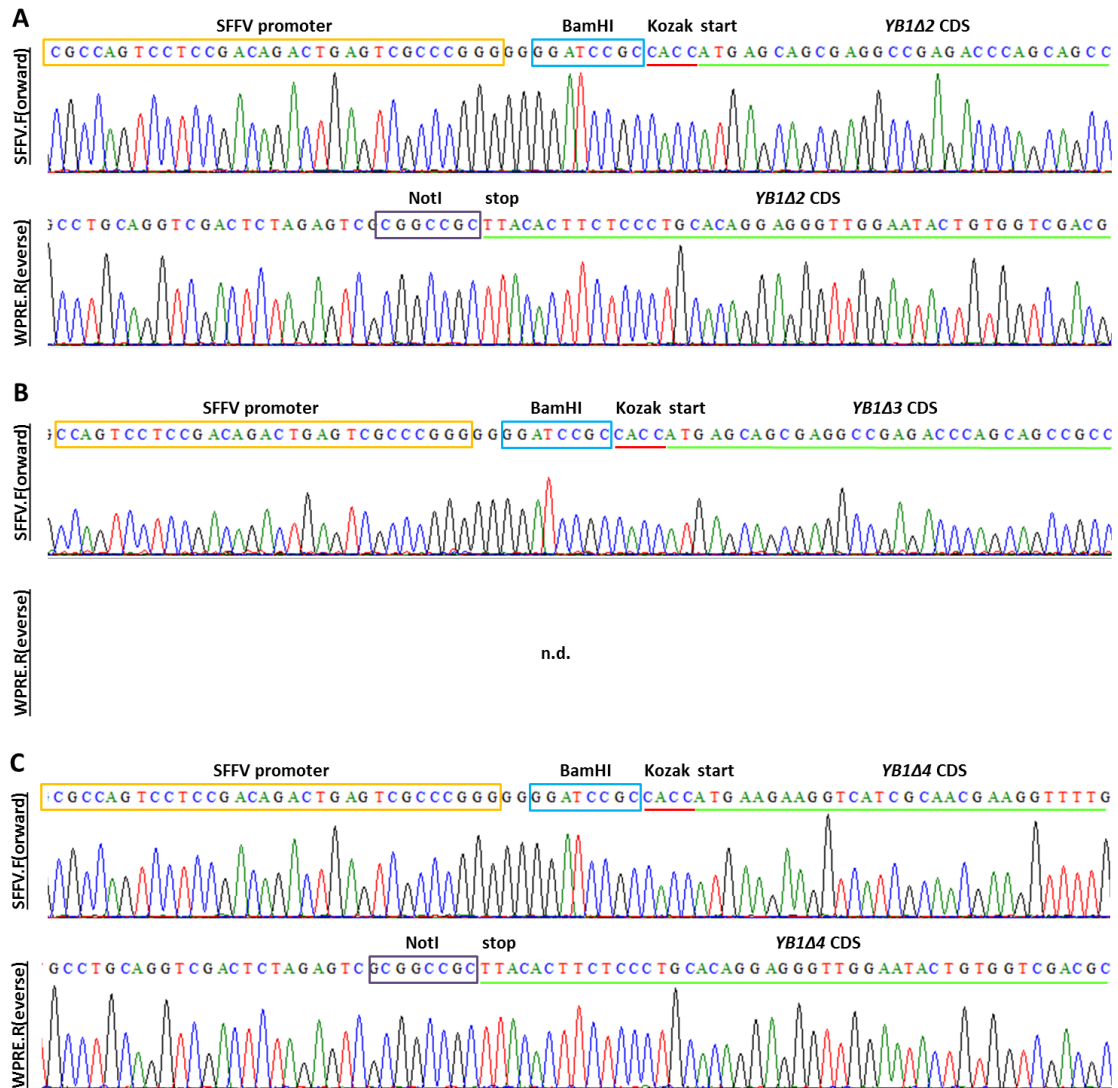


Figure 4.6 Cloning of YB1 full length and truncated YB1-encoding CDS into pDUAL vector backbone for lentiviral vector production. A-C) Sequence and sequence chromatograms demonstrating successful cloning for selected transformants for (A) YB1Δ2, (B) YB1Δ3, and (C) YB1Δ4, into the pDUAL lentiviral transfer vector backbone, using SFFV.F(oward) and WPRE.R(everse) sequencing primers (top and bottom panels, respectively). Coding sequences were considered in-frame with the start codon. SFFV promoter (yellow boxed), BamHI (blue boxed), Kozak sequence (red underlined), coding sequence, green underlined, ATG codon, start, TAA reverse complement codon, stop, and NotI (grey boxed) sequences are annotated. n.d., not determined.

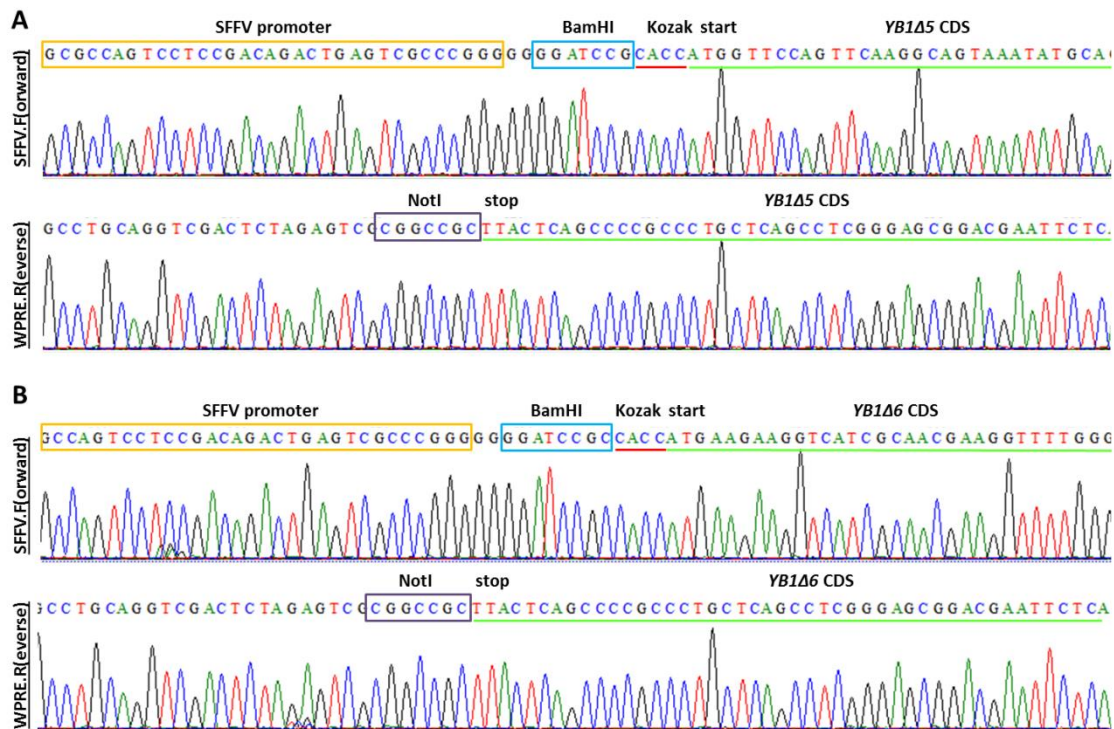


Figure 4.7 Cloning of YBX1 full length and truncated YB1-encoding CDS into pDUAL vector backbone for lentiviral vector production. A-C) Sequence and sequence chromatograms demonstrating successful cloning for selected transformants for (A) YB1Δ5 and (B) YB1Δ6, into the pDUAL lentiviral transfer vector backbone, using SFFV.F(Forward) and WPRE.R(Reverse) sequencing primers (top and bottom panels, respectively). Coding sequences were considered in-frame with the start codon. SFFV promoter (yellow boxed), BamHI (blue boxed), Kozak sequence (red underlined), coding sequence, green underlined, ATG codon, start, TAA reverse complement codon, stop, and NotI (grey boxed) sequences are annotated.

Lentiviral vectors encoding *YB1FL* and *YB1Δ1-Δ6* were stably introduced into YB1 knockout cells. These YB1 knockout cell lines carrying an introduced YB1 mutant were herein referred to as +YB1FL or +YB1Δ1-Δ6, in accordance to the input vector and encoding transgene. The expression and localisation of YB1FL and YB1 mutants were confirmed by ICC and confocal microscopy (Fig. 4.8). Here, either αYB1 (ab114999 or PA5-19453) was used to ensure the spectrum of YB1 truncations could be detected, given epitope targets were known. Fig. 4.8 (panels a-i) demonstrated that negligible fluorescent signal or background fluorescence in 293T cells stained with only the primary αYB1s (panels a-c and d-f for PA5-19453 and ab114999, respectively) or secondary antibodies (panels g-i). 293T cells served as positive control for endogenous YB1 staining, and indeed demonstrated positive YB1 signal, which was predominantly localised in the cytoplasm for either αYB1 antibodies (Fig. 4.8, panels j-l and m-o). Whereas, in contrast, YB1 knockout cells served as negative controls for YB1 staining; and when stained fully with either αYB1 antibodies demonstrated an YB1 knockout phenotype with no detectable fluorescent signal, as previously reported (Fig. 4.8, panels p-r and s-u).

The +YB1FL cell lines recapitulated the expression and localisation of wildtype YB1 (Fig. 4.8, panels v-x, y-aa, and bb-dd). In fact, YB1 retained a strict cytoplasmic localisation even in YB1 knockout cells re-introduced with higher MOIs of lenti-YB1FL vectors (Fig. 4.8, panels y-aa, and bb-dd). When examining +YB1 Δ 1- Δ 6 cell lines, YB1 truncation mutants – YB1 Δ 1, YB1 Δ 3, and YB1 Δ 6 – were characterised as predominantly cytoplasmic in localisation (Fig. 4.8, panels ee-gg, nn-pp, and tt-vv, respectively), and thus displayed similar expression and localisation profiles to parental 293T and wildtype YB1. On the other hand, when compared to the cytoplasmic localisation of 293T control, the YB1 Δ 2 and YB1 Δ 4 truncated mutants exhibited both cytoplasmic and nuclear localisation (Fig. 4.8, panels hh-jj and nn-pp, respectively). Fluorescent signal indicative of YB1 Δ 2 or YB1 Δ 4 protein were thus appreciably detected within the nuclear compartment in these highlighted cells (Fig. 4.8, panel ii or oo; cells are highlighted by white arrows). It was also evident that the YB1 Δ 2 localisation was absent from the nucleolar compartment, as distinguished by high resolution DAPI-staining (Fig. 4.8, panels hh and ii), but difficult to discern whether YB1 Δ 4 protein in +YB1 Δ 4 cell line showed sub-nuclear compartmentalisation. Whereas, YB1 Δ 5 expression and localisation in +YB1 Δ 5 cell line was identified within the nucleus and relatively absent from the cytoplasm. DAPI-staining of the nuclear compartment also helped identify that YB1 Δ 5 signal was particularly detectable in the nucleolar regions of the nuclear compartment (Fig. 4.8, panels qq-ss, see cells highlighted by white arrows). With this in mind, and given the fact that YB1 Δ 5 was a truncation mutant of YB1 Δ 4 and YB1 Δ 6 (of which were predominantly cytoplasmic in localisation), the CRS1 motif was presumably identified as the dominant motif that dictated YB1 localisation to the cytoplasm. This observation was also reflected in Table 4.1 which provides a summary of the observed subcellular localisations for full length YB1 and truncation mutants identified in the present study, and details pertaining to their NLS, CRS, and/or 20S proteasome cleavage site are listed in Table 4.1.

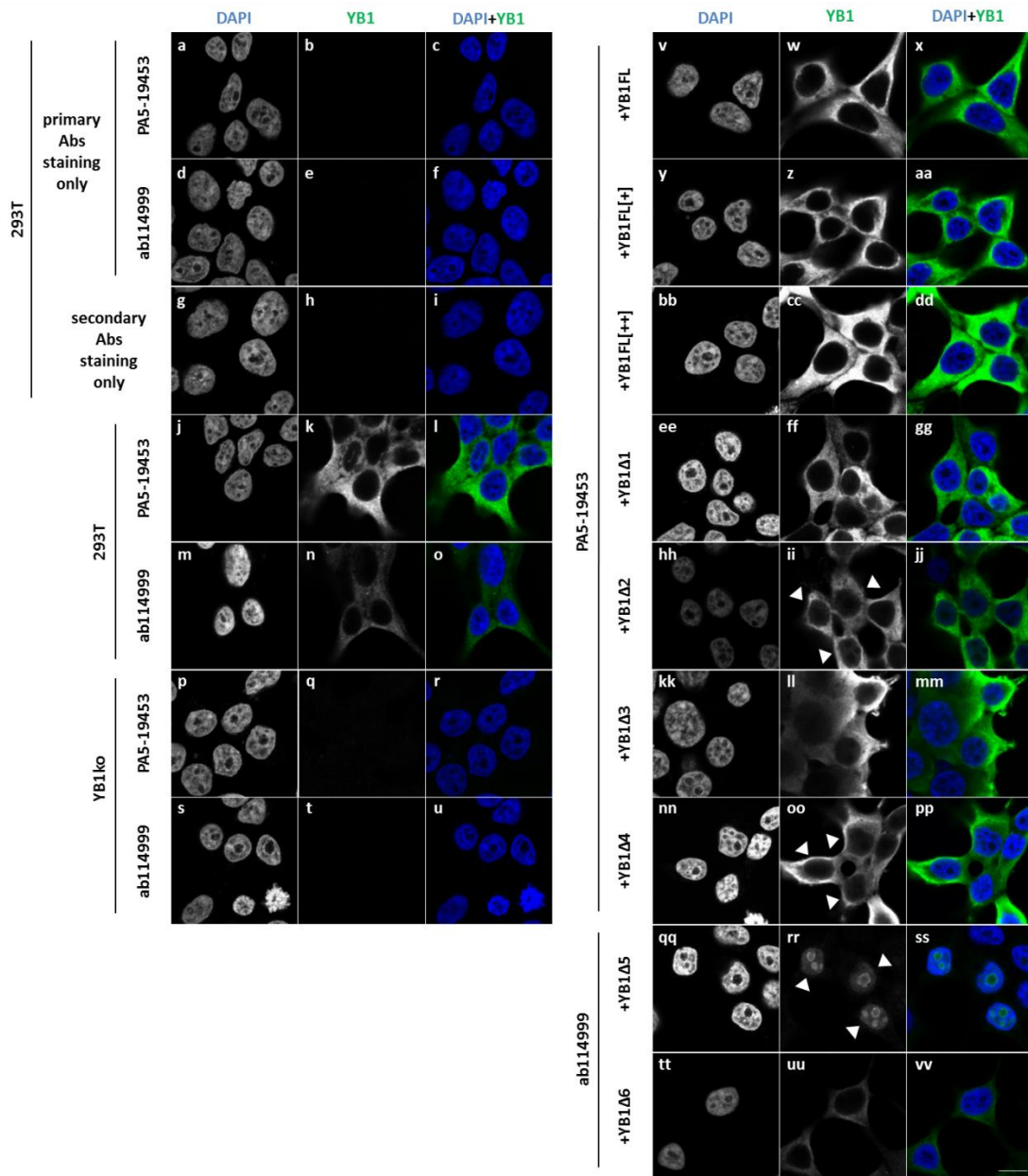


Figure 4.8 ICC and laser scanning confocal microscopy of 293T, YB1 knockout cells, and YB1 knockout cells stably expressing mutant YB1 proteins (YB1FL, YB1Δ1-Δ6). 293T, YB1 knockout, and transduced YB1 knockout cells were fixed, blocked and stained for YB1 and mutant YB1 protein. Staining controls included 293T cells treated with primary Abs only as indicated (**a-c**) and (**d-f**), and secondary Abs only (**g-i**). Positive staining controls included 293T stained with the indicated primary Abs and secondary Ab (**j-i**) and (**m-o**). Negative staining controls included YB1 knockout cells stained with the indicated primary Abs and secondary Abs (**q-r**) and (**s-u**). Mutant YB1 protein expression was detected in transduced YB1 knockout cells for YB1FL (**v-dd**), YB1Δ1 (**ee-gg**), YB1Δ2 (**hh-jj**), YB1Δ3 (**kk-mm**), YB1Δ4 (**nn-pp**), YB1Δ5 (**qq-ss**), and YB1Δ6 (**tt-vv**). The nuclear compartment was distinguished by DAPI-staining. White arrows indicate cells that show mutant YB1 protein localisation in the nucleus, which is in contrast to 293T control. PA5-19453 stains N-terminus of YB1; ab114999 stains C-terminus of YB1. Scale bar = 10μm, and representative for each panel.

Table 4.1 List of YB1FL and YB1 mutants and their subcellular localisation and characteristics

YB1 mutant	aa coverage	localisation	CRS1 (52-101aa)	NLS1 (149-156aa)	NLS2 (185-193aa)	NLS3 (276-292aa)	CRS2 (267-293aa)	20S cleavage site
YB1 (wildtype)	1-324	C	Y	Y	Y	Y	Y	Y
YB1FL	1-324	C	Y	Y	Y	Y	Y	Y
YB1Δ1	1-256	C	Y	Y	Y	-	-	-
YB1Δ2	1-217	C&N	Y	Y	Y	-	-	-
YB1Δ3	1-129	C	Y	-	-	-	-	-
YB1Δ4	52-217	C&N	Y	Y	Y	-	-	-
YB1Δ5	131-324	N	-	Y	Y	Y	Y	Y
YB1Δ6	52-324	C	Y	Y	Y	Y	Y	Y

-, not present; C, cytoplasmic; N, nuclear Y, present

Upon further examination of 293T control, +YB1 Δ 2, +YB1 Δ 4, and +YB1 Δ 5 cell lines for YB1 expression and localisation by post-image analysis, the localisation of YB1 and YB1 mutants were further defined by line-profiling analysis. This involved quantifying the grey intensities (DAPI and YB1 staining) along the length of individual cells (three cells each were selected) via regions of interests (ROIs). Line-profiling through selected cells further confirmed the predominant cytoplasmic localisation of YB1 in 293T control cells (see Fig. 4.9A). Staining of 293T control cells with α YB1 PA5-19453 (Fig. 4.8, panels j-l), or α YB1 ab114999 (Fig. 4.8, panels m-o) antibodies demonstrated a similar predominant cytoplasmic location of YB1 localisation, therefore only the PA5-19453 stained 293T control cells were subjected to line-profiling analysis. In particular, YB1 fluorescent intensity (green line) was primarily evident outside of the nucleus and did not co-localise with DAPI staining (blue line), which differentiated the nuclear compartment of stained cells. This was confirmed by examination of three separate cells by ROI analysis as representatives of the stained cell population (Fig. 4.9A, ROI 1-3).

However, line-profiles of +YB1 Δ 2 cells defined YB1 Δ 2 (green line) localisation throughout the entire length of stained cells, and included localisation with DAPI-staining (blue line) as a differentiating stain for the nuclear compartment (Fig. 4.9B). Given ROI 1-3 were each drawn through the length of stained cells and included nucleolar compartments, line-profiling indicated that YB1 Δ 2 fluorescence did not coincide with nucleolar compartments well. This was indicated by sharp troughs of YB1 Δ 2 (green line) and DAPI (blue line) fluorescent intensities, at the same points along the length of the ROI 1-3. Stained +YB1 Δ 4 cells showed similar line-profiles for YB1 Δ 4 to +YB1 Δ 2 stained cells (Fig. 4.9C). Except, however, line-profiling of YB1 Δ 4 signal was picked up throughout the nucleus and not compartmentalised strictly in the nucleoplasm as was observed for YB1 Δ 2. Stained +YB1 Δ 5 cell line, on the other hand, exhibited YB1 Δ 5 (green line) localisation strictly in the nucleus as per DAPI-staining (blue line) (Fig. 4.9D). Furthermore, the highest fluorescent intensities of YB1 Δ 5 was principally detected at the nucleolar compartment by line-profiling, where peak YB1 Δ 5 fluorescent intensities coincided or correlated with troughs in DAPI-fluorescent intensities.

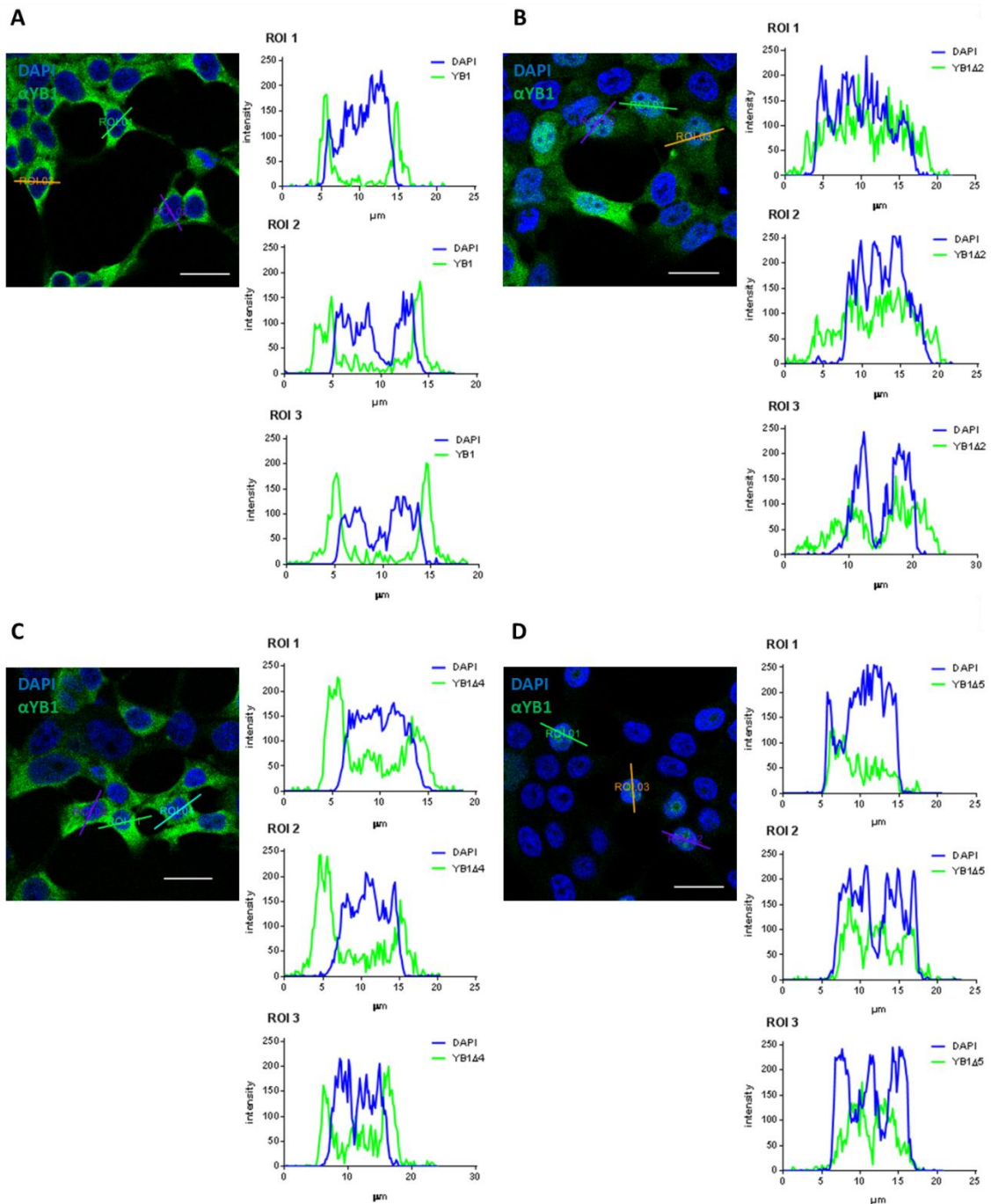


Figure 4.9 Mutant YB1 protein (YB1Δ2, YB1Δ4 and YB1Δ5) show nuclear localisation by line-profiling. Line graphs represent line-profiling of selected ROIs from 3 separate cells (pictured) of **(A)** 293T (control), or YB1ko cells stably expressing **(B)** YB1Δ2 and **(C)** YB1Δ4, and **(D)** YB1Δ5, using LAS X image analysis software. The grey intensities representing YB1 or mutant YB1 (green line) was spatially related to the nuclear compartment distinguished by DAPI-staining (blue line). The subcellular distribution of YB1Δ2 and YB1Δ5 mutants was therefore distinguished as diffuse throughout the cell (but absent from nucleolar regions), and considerably nucleolar localised for YB1Δ2 and YB1Δ5 mutants, respectively. Scale = 20μm. ROI, region of interest.

To rule out z-plane discrepancy in the localisation of YB1Δ2, YB1Δ4, and YB1Δ5 truncated mutants observed by ICC and line-profiling, orthogonal cross-sections were examined, and compared against wildtype YB1 in 293T positive control cell line. Altogether, orthogonal cross-

sections further confirmed the localisations of YB1, YB1 Δ 2, YB1 Δ 4, and YB1 Δ 5 protein as was previously reported (Fig. 4.8) and line-profiling (Fig. 4.9). While 293T control cells displayed wildtype YB1 distributed throughout the cell cytoplasm, and relatively absent from the nuclei (Fig. 4.10A), YB1 Δ 2 and YB1 Δ 4 protein localisation were found throughout the length (x, y planes) and depth (z-plane) of +YB1 Δ 2 (Fig. 4.10B) and +YB1 Δ 4 (Fig. 4.10C) stained cells. Although, YB1 Δ 2 and YB1 Δ 4 staining was primarily detected in the cell cytoplasm, notable fluorescent signal is also picked up and overlaps with the nuclear compartment throughout the x-, y- and z- planes. On the other hand, YB1 Δ 5 protein was reaffirmed as quite strictly nuclear in localisation, but further showed a rather defined compartmentalisation to nucleoli regions (Fig. 4.10D), as distinguished by high resolution DAPI-staining.

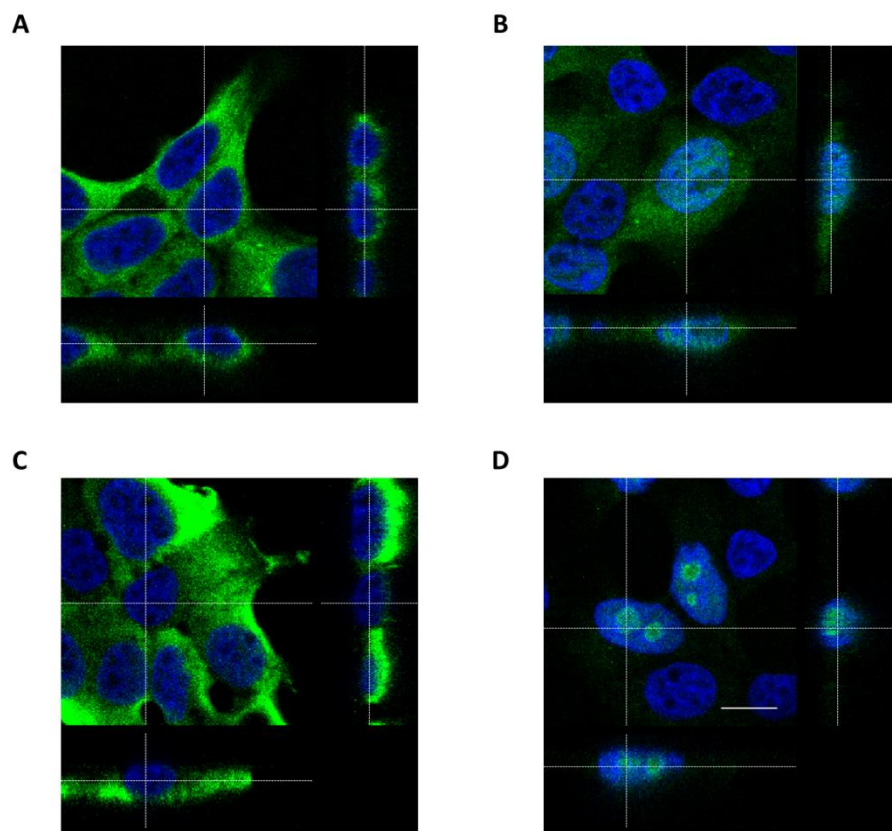


Figure 4.10 Orthogonal cross-sections further identify YB1 Δ 2, YB1 Δ 4, and YB1 Δ 5 protein localisations. Subcellular localisation of wildtype YB1, and truncated YB1 Δ 2, YB1 Δ 4, and YB1 Δ 5 were further defined and confirmed by examining orthogonal cross-sections derived from z-stacks of cells stained with α YB1 (green) and DAPI (blue). Therefore, a strictly cytoplasmic localisation was identified for **(A)** wildtype YB1 using 293T positive control; a predominant cytoplasmic with notable nuclear localisation for **(B)** YB1 Δ 2 using +YB1 Δ 2 and **(C)** YB1 Δ 4 using +YB1 Δ 4 cells. Finally, a nucleoli-defined localisation for **(D)** YB1 Δ 5 was identified after staining +YB1 Δ 5 cells.

Finally, the extent of YB1, YB1 Δ 2, YB1 Δ 4, and YB1 Δ 5 staining in the nucleus was quantified by calculating average nuclear:cytoplasmic ratios (N/C). In summary, significant localisation of YB1 Δ 2, YB1 Δ 4 and YB1 Δ 5 in the nuclear compartment was quantified when compared to 293T

control. 293T control cells demonstrated a clear cytoplasmic enrichment of wildtype YB1 when staining cells with either α YB1 – PA5-19453 or ab114999, showing N/C ratio of -1.283 ± 0.184 and -1.392 ± 0.247 , respectively after log transformation (Fig. 4.11). Similarly, but to a significantly reduced level compared to 293T and wildtype YB1, YB1 Δ 4 also presented with cytoplasmic enrichment with N/C ratio of -0.293 ± 0.271 (Fig. 4.11; $n = 30-40$ cells analysed, $P < 0.0001$). This indicated that YB1 Δ 4 was predominantly expressed in the cytoplasm, but the reduced N/C ratio compared to that quantified in 293T controls suggests that YB1 Δ 4 was also quantified in the nuclear compartment. In contrast to this, YB1 Δ 2 and YB1 Δ 5 proteins, displayed significant nuclear staining with N/C ratios of 0.443 ± 0.28 and 0.971 ± 0.35 after log transformation compared to 293T control with their corresponding α YB1 antibodies (Fig. 4.11; $n = 30-40$ cells each cell analysed, $P < 0.0001$). Therefore, a particular shift in nuclear staining was quantified for YB1 Δ 4, YB1 Δ 2, and YB1 Δ 5, with nuclear predominance quantified for the latter two.

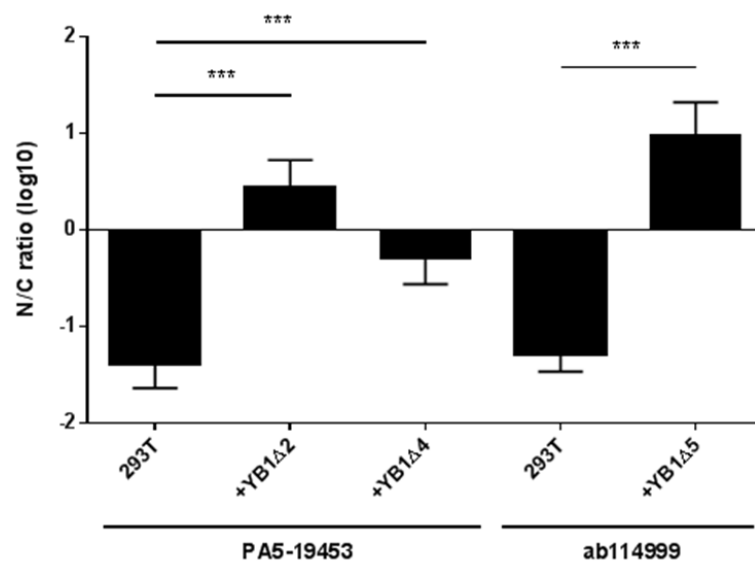


Figure 4.11 Nuclear/cytoplasmic (N/C) ratios of YB1 Δ 2, YB1 Δ 4, and YB1 Δ 5 compared to 293T control cell line. Average N/C ratios (after log₁₀ transformation) of detected YB1 signal in nucleus and cytoplasm were calculated from between 30-40 individual cells, for each cell line that demonstrated nuclear localisation of truncated YB1 in YB1 Δ 2-, YB1 Δ 4-, and YB1 Δ 5- expressing cell lines. This was compared to the N/C ratio of detected, wildtype YB1 in 293T for the corresponding α YB1 (PA5-19453 or ab114999). Error bars reflect \pm SD from mean, $n = 30-40$ cells analysed for each cell line; ***, $P < 0.0001$.

4.2.4 Expression of YB1 and select truncated mutants rescues the cytotoxicity phenotype in YB1 knockout cell lines

We previously described that the YB1 knockout phenotype was significantly sensitive to cytotoxicity after triple transfection using Calcium Phosphate precipitation with chloroquine and with chloroquine alone (Section 3.2.11). We then endeavoured to explore whether or not

this phenotype could be rescued, and which functional domain of YB1 attributed to this undesirable phenotype using LDH cytotoxicity assays. Firstly, all non-treated mutant cell lines (+YB1FL and +YB1 Δ 1- Δ 6) demonstrated low %cytotoxicity similar to 293T (8.75 \pm 1.38%) and YB1 knockout (11.68 \pm 1.09%) controls (n = 3 each cell line; $P > 0.05$) (Figs. 4.12, 4.13-14A).

Significant increase in %cytotoxicity was observed in YB1 knockout cells (62.14 \pm 0.23%) compared to control 293T (22.97 \pm 1.33%) 48h post-treatment with chloroquine (Fig. 4.12). However, we observed a rescue in the adverse sensitivity to chloroquine when full length YB1 or select YB1 mutants were expressed in an YB1 knockout background (Fig. 4.12). Stable expression of full length YB1 significantly reduced the %cytotoxicity after addition of chloroquine to 33.02 \pm 2.27%, 29.6 \pm 0.94%, and 32.63 \pm 2.67% for +YB1FL, +YB1FL[+], and +YB1FL[++] , respectively (Fig. 4.12; n = 3 each cell line; $P < 0.0001$) compared to YB1 knockout cells. These %cytotoxicities were appreciably close to that exhibited by 293T control (22.97 \pm 1.33%), of which difference to 293T control remained statistically significant (n = 3 each cell line; $P < 0.0001$, $P < 0.05$, and $P < 0.0001$, respectively for and +YB1FL, +YB1FL[+], and +YB1FL[++]). This indicated that a complete rescue was not fully achieved by reintroducing YB1 expression.

Stable expression of particular YB1 mutants also significantly reduced the cell cytotoxicity 48h post-chloroquine treatment. The +YB1 Δ 2, +YB1 Δ 3, and +YB1 Δ 4 cells showed the most impressive reduction of %cytotoxicity when compared to 293T or YB1 knockout cells after chloroquine treatment, exhibiting average %cytotoxicity of 9.68 \pm 1.76%, 7.82 \pm 1.86%, and 22.44 \pm 3.28%, respectively (Fig. 4.12; compare YB1 knockout to +YB1 Δ 2 and +YB1 Δ 3 after chloroquine treatment; n = 3 each cell line, $P < 0.0001$). The +YB1 Δ 4 cell line also demonstrated a significant reduction in %cytotoxicity when compared to YB1 knockout (n = 3 each cell line; $P < 0.0001$), and was considered comparable to 293T control, indicating that stable YB1 Δ 4 expression was sufficient to alleviate the cell cytotoxic effects of chloroquine to wildtype levels (Fig. 4.12; compare 293T to +YB1 Δ 4 after chloroquine treatment). Significant reduction in cell cytotoxicity was also achieved by +YB1 Δ 1 (41.74 \pm 3.79%) and +YB1 Δ 5 (37.5 \pm 4.81%) cells lines when compared to YB1 knockout, 48h after the addition of chloroquine (Fig. 4.12; n = 3 each cell line, $P < 0.0001$). The %cytotoxicity in +YB1 Δ 1 or +YB1 Δ 5 cells was still significantly higher than that of control 293T cells (Fig. 4.12; n = 3 each cell line, $P < 0.0001$), indicating a partial rescue by YB1 Δ 1 or YB1 Δ 5 expression. Mutant +YB1 Δ 6 cells

exhibited no rescue with elevated %cytotoxicity ($71.52 \pm 1.12\%$) when compared to 293T or YB1 knockout (Fig. 4.12A; $n = 3$ each cell line, $P < 0.0001$).

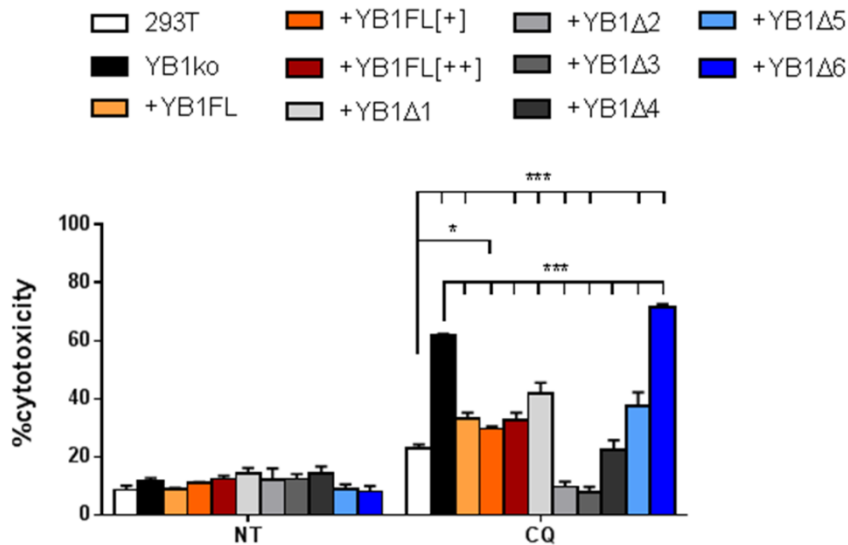


Figure 4.12 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h post-treatment with chloroquine. Results presented indicated that YB1 knockout cells were significantly susceptible to cell cytotoxicity after chloroquine treatment, but cell cytotoxicity was rescued or partially rescued by the stable expression of YB1FL or select YB1 truncated mutants. LDH cytotoxicity profiles were calculated for 293T (solid white bars) and YB1 knockout (solid black bars), +YB1FL (solid yellow bars), +YB1FL[10] (solid orange bars), +YB1FL[300] (solid red bars), +YB1Δ1 (solid off white bars), +YB1Δ2 (solid light grey bars), +YB1Δ3 (solid medium grey bars), +YB1Δ4 (solid dark grey bars), +YB1Δ5 (solid light blue bars), and +YB1Δ6 (solid dark blue bars), after treatment with chloroquine (final $25\mu\text{M}$) 48h post-treatment. LDH activity for each cell line and treatment was calculated as a % of max lysis of each respective non-treated control cell line. Error bars reflect $\pm\text{SD}$ from mean, $n = 3$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.0001$. CQ, chloroquine; NT, no treatment; YB1ko, YB1 knockout.

A similar pattern of %cytotoxicity was observed for 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h after triple transfection using the Calcium Phosphate precipitation method supplemented with chloroquine (Fig. 4.13A). Averages of $27.45 \pm 1.77\%$ and $63.91 \pm 1.35\%$ cytotoxicity were measured for 293T and YB1 knockout cell lines after triple transfection with chloroquine, respectively; and difference considered statistically significant ($n = 3$ each cell line, $P < 0.0001$). The stable expression of full length YB1 in YB1 knockout cell lines resulted in the rescue of the cell cytotoxic phenotype to wildtype levels in +YB1FL cell lines – averages of $29.56 \pm 2.8\%$, $25.99 \pm 1.04\%$, and $29.9 \pm 1.84\%$ cytotoxicity was demonstrated for +YB1FL, +YB1FL[+], and +YB1FL[++] cell lines, respectively. These reductions in average %cytotoxicity were statistically significant when compared to YB1 knockout cells only (Fig. 4.13A; $n = 3$ each cell line, $P < 0.0001$), but comparable to 293T control.

Similarly, average %cytotoxicity were rescued to $19.08 \pm 4.71\%$, $14.48 \pm 0.93\%$, $12.63 \pm 1.87\%$, and $35.43 \pm 0.95\%$ for +YB1 Δ 2, +YB1 Δ 3, +YB1 Δ 4 and +YB1 Δ 5, respectively, after triple transfection with chloroquine. These differences of which were considered statistically significant when compared to YB1 knockout or 293T control (Fig. 4.13A; $n = 3$ each cell line, $P < 0.0001$ for either comparisons, $P < 0.01$ between 293T and +YB1 Δ 5). Therefore, stable YB1 Δ 5 expression was identified to partially rescue the cytotoxic effects of triple transfection supplemented with chloroquine. Finally, a lack of rescue of the cytotoxic effects of triple transfection supplemented with chloroquine was observed for +YB1 Δ 1 and +YB1 Δ 6 (Fig. 4.13A), which correlated with cell cytotoxicity observations 48h post-treatment with chloroquine only (Fig. 4.12). More specifically, average %cytotoxicities of $62.67 \pm 0.41\%$ and $70.01 \pm 4.66\%$ were calculated for +YB1 Δ 1 and +YB1 Δ 6, respectively. When compared to 293T control, the elevated %cytotoxicity showed statistical significance (Fig. 4.13A; $n = 3$ each cell line, $P < 0.0001$). Whereas, comparing to YB1 knockout, %cytotoxicity exhibited by +YB1 Δ 1 was considered comparable, but modestly statistically significance when compared to +YB1 Δ 6 (Fig. 4.13A; $n = 3$ each cell line; $P < 0.05$). The transfection efficiency of pAAV-hrGFP plasmid was also considered, and reported as comparable (>97%) between 293T control, YB1 knockout, +YB1FL, and +YB1 Δ 1- Δ 6 cell lines, 48h post-transfection using Calcium Phosphate precipitation supplemented with chloroquine (Fig. 4.13B).

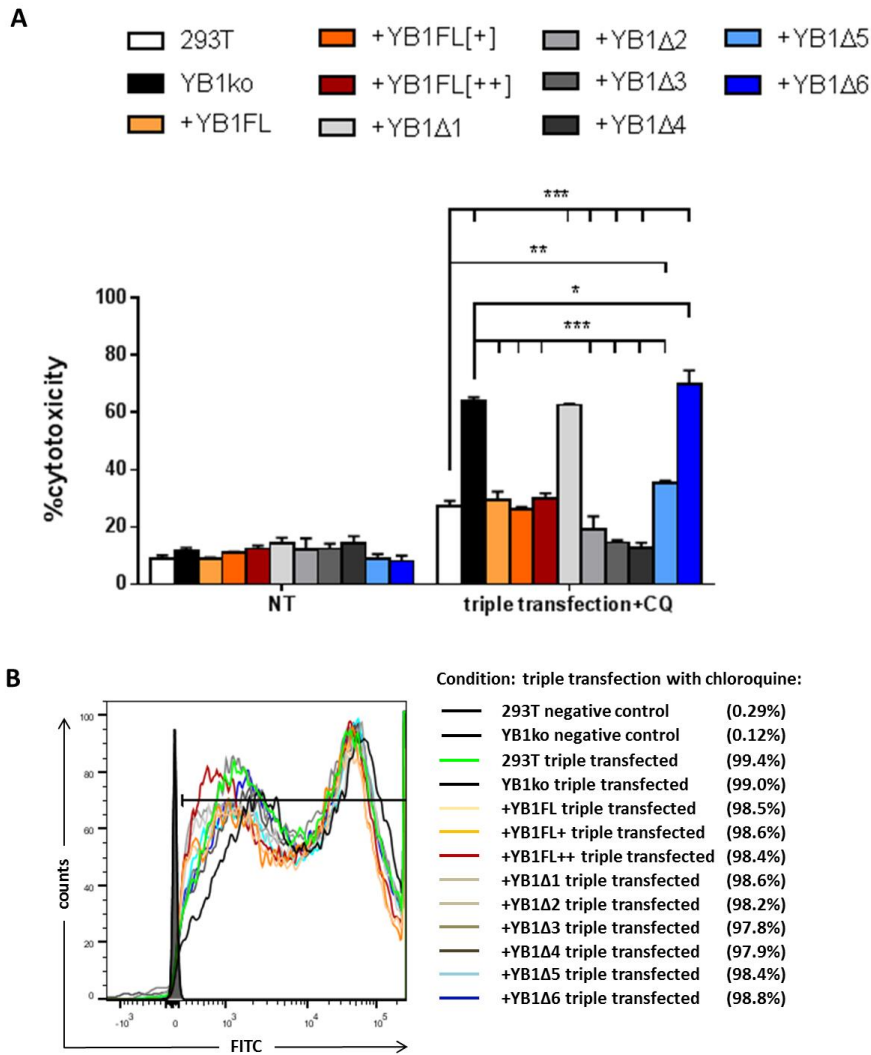


Figure 4.13 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h post-transfection using Calcium Phosphate precipitation method supplemented with chloroquine. A) LDH cytotoxicity profiles were calculated for 293T (solid white bars) and YB1 knockout (solid black bars), +YB1FL (solid yellow bars), +YB1FL[+] (solid orange bars), +YB1FL[++] (solid red bars), +YB1Δ1 (solid off white bars), +YB1Δ2 (solid light grey bars), +YB1Δ3 (solid medium grey bars), +YB1Δ4 (solid dark grey bars), +YB1Δ5 (solid light blue bars), and +YB1Δ6 (solid dark blue bars), 48h post-transfection with chloroquine (final 25μM) for rAAV2 vector production. LDH activity for each cell line and treatment was calculated as a % of max lysis of each respective non-treated control cell line. Results presented indicated that YB1 knockout cells were significantly susceptible to cell cytotoxicity after chloroquine treatment, but cell cytotoxicity was rescued or partially rescued by the stable expression of YB1FL or select YB1 truncated mutants. Error bars reflect \pm SD from mean, $n = 3$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.0001$. **B)** Comparable transfection efficiencies between 293T, YB1 knockout, YB1FL, and YB1Δ1-Δ6 cell lines were measured by flow cytometry, 48h post-transfection for rAAV2 vector production. This was compared to negligible fluorescence of non-transfected 293T and YB1 knockout cells as negative controls. Transfection efficiencies are presented as % in parentheses. CQ, chloroquine; NT, non-treated; YB1ko, YB1 knockout.

Cells triple transfected without chloroquine were also examined. The cytotoxicity profiles between 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines, further substantiated that chloroquine was the predominant effector that contributed to the adverse cell cytotoxicity in YB1 knockout

cells (Fig. 4.14A). Calculated %cytotoxicities after triple transfection without chloroquine were considered low and relatively similar to %cytotoxicities exhibited by non-treated 293T, YB1 knockout, and +YB1 Δ 1- Δ 5 cell lines (Fig. 4.14A). No significant difference in %cytotoxicity was measured between control 293T (12.06 \pm 0.69%) and remainder cell lines after triple transfection without chloroquine. However, +YB1 Δ 6 exhibited significant cell cytotoxicity (43.64 \pm 3.48%) when compared against 293T control or YB1 knockout – average of 13.23 \pm 0.31% (Fig. 4.13A; n = 3 each cell line, $P < 0.0001$ for each comparison). Additionally, a significant reduction in %cytotoxicity was also observed for +YB1 Δ 5 cell line (7.36 \pm 1.67%), when compared to YB1 knockout cell line – 13.23 \pm 0.31% (Fig. 4.13A; n = 3 each cell line, $P < 0.05$). These findings were further appreciated by comparable transfection efficiencies (>93%) for the pAAV2-hrGFP plasmid of triple transfected cells measured 48h post-transfection, by flow cytometry (Fig. 4.14B).

Therefore, from LDH cytotoxicity data, it was generally considered that triple transfection using the Calcium Phosphate precipitation methodology without the supplementation of chloroquine did not contribute to the cytotoxic phenotype in YB1 knockout or +YB1 Δ 1- Δ 5 cell lines. In fact, these data imply that chloroquine reagent remains the predominant contributor and effector of cell cytotoxicity in YB1 knockout cells, as previously reported (Chaper 3, 3.3.11). The cytotoxic effect of chloroquine could be rescued in the present study – fully or partially – by restoring the stable expression of full length YB1 and YB1 Δ 5. Moreover, the cytotoxic effect of chloroquine was further reduced to below the baseline reference by the stable expression of YB1 Δ 2- Δ 4. It was noted that YB1 Δ 6 was unable to rescue YB1 knockout cells from chloroquine-induced cytotoxicity, which may be a consequence of the complete truncation of the A/P domain. Nonetheless, the data thus far indicates a general consensus that the CSD of wildtype YB1 protein exerts a dominant protective function, which may be negatively regulated in part by the A/P domain, but more-so when the entire CTD of YB1 is expressed in tandem with its CSD. Therefore, a complex interplay between YB1's A/P, CSD and CTD domains is rather integral to facilitating resistance to the cytotoxic effects of chloroquine.

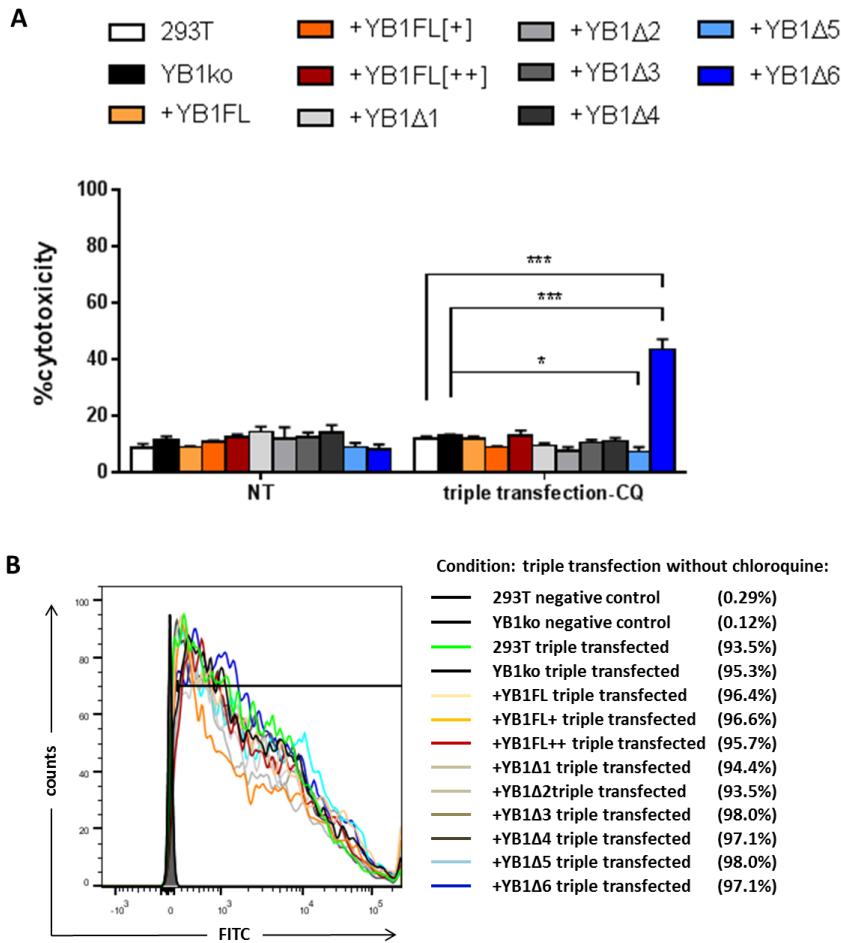
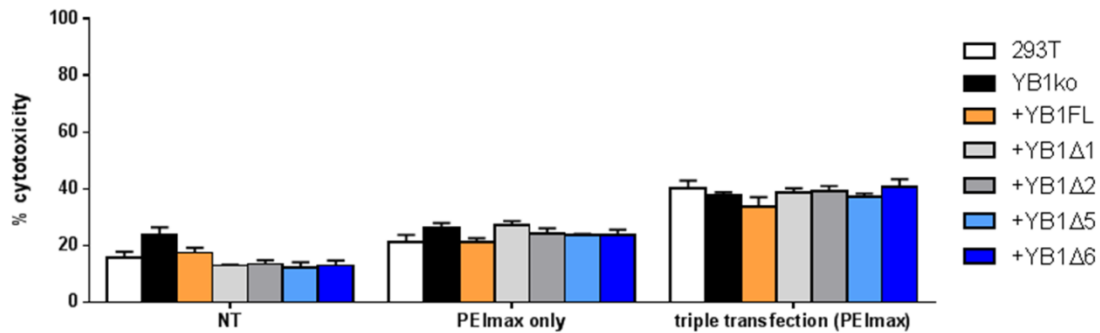


Figure 4.14 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 48h post-transfection using Calcium Phosphate precipitation method without chloroquine. A) LDH cytotoxicity profiles were calculated for 293T (solid white bars) and YB1 knockout (solid black bars), +YB1FL (solid yellow bars), +YB1FL[+] (solid orange bars), +YB1FL[++] (solid red bars), +YB1Δ1 (solid off white bars), +YB1Δ2 (solid light grey bars), +YB1Δ3 (solid medium grey bars), +YB1Δ4 (solid dark grey bars), +YB1Δ5 (solid light blue bars), and +YB1Δ6 (solid dark blue bars), 48h post-transfection for rAAV2 vector production. Results presented indicated that generally cell cytotoxicity was comparable between +YB1FL, YB1Δ1-Δ5 and control 293T and YB1 knockout cells after triple transfection for rAAV2 vector production without chloroquine. However, +YB1Δ6 showed a significant increase in %cytotoxicity after triple transfection for rAAV2 vector production. LDH activity for each cell line and treatment was calculated as a % of max lysis of each respective non-treated control cell line. Error bars reflect \pm SD from mean, $n = 3$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.0001$. **B)** Comparable transfection efficiencies between 293T, YB1 knockout, YB1FL, and YB1Δ1-Δ6 cell lines were measured by flow cytometry, 48h post-transfection for rAAV2 vector production. This was compared to negligible fluorescence of non-transfected 293T and YB1 knockout cells as negative controls. Transfection efficiencies are presented as % in parentheses. NT, no treatment. CQ, chloroquine; NT, no treatment; YB1ko, YB1 knockout.

The effect of stable expression of full length YB1 and YB1 truncated mutants (YB1Δ1-Δ6) on cell cytotoxicity was also explored 72h post-triple transfection using PEI_{max}. To summarise, comparable %cytotoxicities were measured between 293T control, YB1 knockout, and +YB1FL, +YB1Δ1, +YB1Δ2, +YB1Δ5 and +YB1Δ6 cell lines, for each respective treatment/transfection (Fig. 4.13A). In fact, +YB1FL, +YB1Δ1, +YB1Δ2, +YB1Δ5 and +YB1Δ6 cells did not exhibit

%cytotoxicities that exceeded that for 293T ($15.89\pm 1.99\%$ and $21.44\pm 2.38\%$) or YB1 knockout ($23.87\pm 2.6\%$ and $26.52\pm 1.43\%$) controls, 72h after no-treatment or treatment with PEI_{max} only, respectively. Similarly, comparable %cytotoxicities was measured 72h post-triple transfection using PEI_{max} for 293T ($40.29\pm 2.73\%$), YB1 knockout ($37.94\pm 0.88\%$), +YB1FL ($33.65\pm 3.4\%$), +YB1Δ1 ($38.58\pm 1.66\%$), +YB1Δ2 (39.23 ± 1.78), +YB1Δ5 (37.28 ± 1.78), and +YB1Δ6 ($40.92\pm 2.48\%$). These findings were further appreciated by comparable transfection efficiencies (>94%) was achieved across all cell lines used 72h post-triple transfection using PEI_{max} (Fig. 4.13B). Overall, LDH cytotoxicity data indicated that YB1 knockout, and rescue of full length YB1 or mutant YB1 expression did not contribute to a cell cytotoxic phenotype after exposure to PEI_{max} reagent or triple transfection for rAAV2GFP vector production using PEI_{max}.

A



B

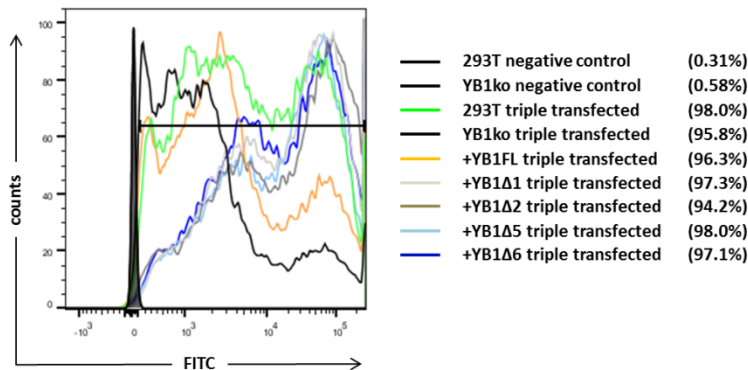


Figure 4.15 LDH cytotoxicity measurements from 293T, YB1 knockout, and +YB1Δ1-Δ6 cell lines 72h post-transfection with PEI_{max}. **A)** LDH cytotoxicity profiles were calculated for 293T (solid white bars) and YB1 knockout (solid black bars), +YB1FL (solid yellow bars), +YB1Δ1 (solid light grey bars), +YB1Δ2 (solid grey bars), +YB1Δ5 (solid light blue bars), and +YB1Δ6 (solid dark blue bars), after no-treatment, treatment with PEI_{max} reagent only, or triple transfection for rAAV2 vector production using PEI_{max}, 72h post-treatment. The %cytotoxicity was considered comparable between all cell lines despite transfection with PEI_{max} reagent. LDH activity for each cell line and treatment was calculated as a % of max lysis of each respective no-treatment control cell line. Error bars reflect \pm SD from mean, n = 3 each cell line. **B)** Flow cytometry analysis for transfection efficiencies for the pAAV2-hrGFP reporter plasmid 72h post-transfection of 293T, YB1 knockout, +YB1FL, and +YB1Δ1-Δ6 cell lines, against 293T and YB1 knockout non-treated negative controls. Transfection efficiencies are presented as % in parentheses. NT, no-treatment.

4.2.5 Examination potential interactions between YB1 and AAV during vector production

In order to accurately visualise the localisation of rAAV2 vector, Rep, Cap or YB1 proteins by ICC, the pAAV2-FLuc vector was designed and used to offset the green-fluorescence signal produced by GFP expressed from the pAAV2-hrGFP expression vector (Fig. 4.16). Figure 4.16A shows the strategy employed to clone a *FLuc* transgene fragment from pJET1.2-FLuc into the pAAV2-MCS expression vector. Complete digestion of the pAAV2-MCS expression vector (Fig. 4.16B) and the *FLuc* transgene fragment dropout (approximately 1700bp dropout) from pJET1.2-FLuc (Fig. 4.16C-D) were ligated to generate a new vector – pAAV2-FLuc, and confirmed by analytical restriction enzyme digestion for the *FLuc* transgene dropout (Fig. 4.16E).

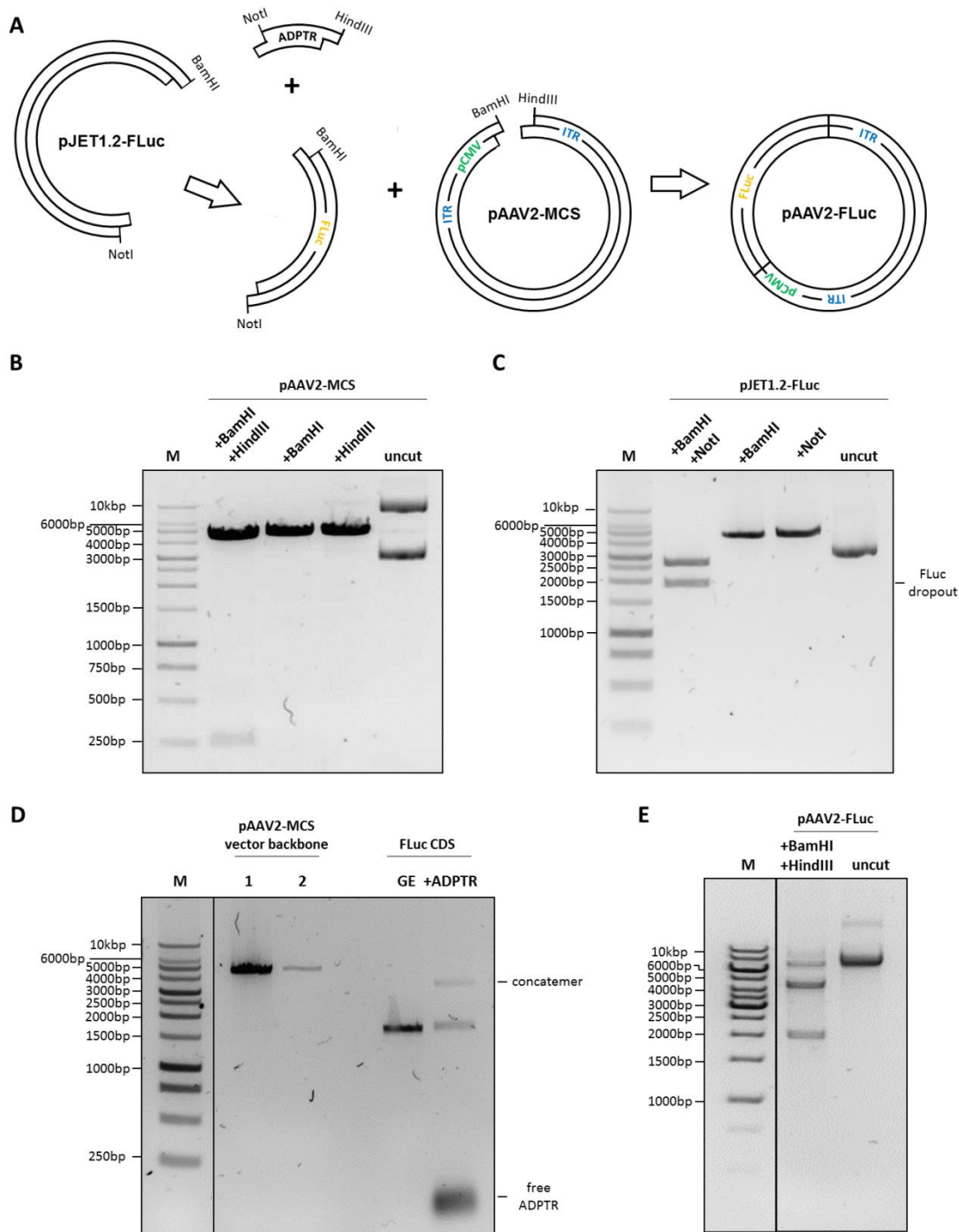


Figure 4.16 Cloning to generate pAAV2-FLuc plasmid for rAAV2FLuc vector production. **A)** Schematic the cloning strategy used to generate pAAV2-FLuc plasmid, from pAAV2-MCS vector and pJET1.2-FLuc, which harbours the *FLuc* transgene. **B)** Double restriction enzyme digestion of pAAV2-MCS expression vector was distinguished from single and uncut controls by 1% TAE agarose gel electrophoresis. **C)** Double restriction enzyme digestion of pJET1.2-FLuc demonstrated *FLuc* transgene dropout (approximately 1700bp) by 1% agarose gel electrophoresis, especially when compared to single and uncut plasmid controls. **D)** 1% agarose gel electrophoresis of gel purified pAAV2-MCS vector backbone (lane annotated 1), which was also subjected to double restriction enzyme digestion (lane annotated 2), gel purified *FLuc* transgene (lane GE). Finally *FLuc* was pre-ligated with the NotI-HindIII ADPTR molecule (lane +ADPTR). Cloning of *FLuc* into pAAV2-MCS was verified by **(E)** analytical restriction enzyme digestion and 1% TAE agarose gel electrophoresis (approximately 1700bp dropout), compared to uncut control. ADPTR, NotI-HindIII adaptor molecule; *FLuc*, firefly luciferase transgene; GE, gel extracted; M, 1kb GeneRuler DNA Ladder.

In order to identify whether or not the YB1 knockout phenotype influenced the localisation of rAAV2 vector particle, AAV2 Cap or Rep proteins, we examined the co-localisation between YB1 or YB1 Δ 1- Δ 6, with rAAV2 vector particle (Fig. 4.18), or AAV2 Cap (Fig. 4.19), or AAV2 Rep (Fig. 4.20) proteins. Firstly, background fluorescence was determined as absent for triple transfected 293T cells stained with primary or secondary antibodies, as shown (Fig. 4.17). The influence of selected functional domains of YB1 on the localisation of the aforementioned AAV2 proteins was also examined in parallel using the +YB1FL and +YB1 Δ 1- Δ 6 cell lines, that stably expressed full length and YB1 mutants.

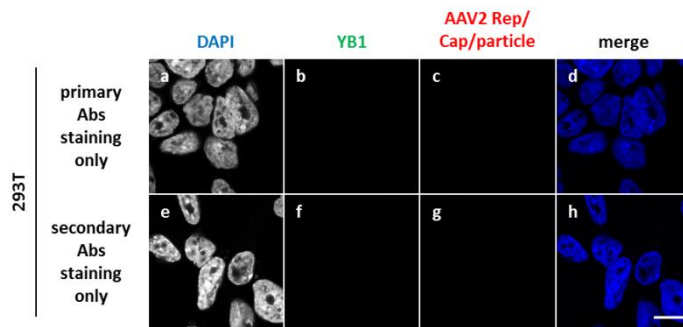


Figure 4.17 ICC for background fluorescence of triple transfected 293T control cells after staining with primary or secondary antibodies. 293T control cells were triple transfected for rAAV2FLuc vector production, and 72h post-transfection cells were fixed, blocked and then stained with either primary α YB1 (PA5-19543 and ab114999), A20 (AAV2 particle), 10R-A140A (AAV2 Rep), and B1 (AAV2 Cap), or secondary Abs (α Rabbit Alexa Fluor 488 and α Mouse DyLight 633). Control staining showed that only background fluorescence or the absence of fluorescent after staining with primary α YB1 (**a-d**) or secondary Abs (**e-h**), was exhibited in either laser channels. DAPI staining served to differentiate the nuclear compartment. Scale bar = 10 μ m, and representative for all panels. Abs, antibody.

Triple transfected 293T control cells demonstrated a predominant cytoplasmic localisation of wildtype YB1 (green fluorescence), with no signal detected in the nuclear compartment when using either α YB1 to target YB1's N- (Figs. 4.18-4.20, panels b and d), or C- terminus (Figs. 4.18-4.20, panels f and h), confirming early observations shown in Fig. 4.8. The rAAV2 particle in triple transfected 293T control cells exhibited a nuclear localisation with notable preference to the nucleolar regions (Fig. 4.18, panels c and d), which was in general agreement with previous studies (Wistuba et al, 1997; Sonntag et al, 2017). Therefore, no co-localisation between wildtype YB1 and rAAV2 particle was inferred from the current study. No fluorescent signal corresponding to YB1 protein was detected in triple transfected YB1 knockout cells (Figs. 4.18-4.20, panels j and l), whereas, rAAV2 particle was detected in YB1 knockout cells with similar nuclear localisation (Fig. 4.18, panels k and l) as triple transfected 293T control. The localisation of YB1FL protein in +YBFL, +YBFL[+], and +YBFL[++] and mutants YB1 Δ 1, Δ 3, and Δ 6 was predominantly cytoplasmic and as previously described in section 4.3.3. In short, YB1FL (Figs. 4.18-4.20, panels j and l, n and o, and r and t), YB1 Δ 1 (Figs. 4.18-4.20, panels v and x),

YB1 Δ 3 (Figs. 4.18-4.20, panels dd and ff), and YB1 Δ 6 (Figs. 4.18-4.20, panels hh and jj) maintained their predominantly cytoplasmic localisation. YB1 Δ 2 and YB1 Δ 4 proteins showed diffuse localisation throughout their respective cell lines (Figs. 4.18-4.20, panels z and bb and panels hh and jj, respectively), with YB1 Δ 2 showing some exclusion from nucleolar regions, as previously observed. Finally, YB1 Δ 5 protein consistently demonstrated a predominantly nucleolar localisation (Figs. 4.18-4.20, panels ll and nn) as previously identified.

The localisation of rAAV2 particles was within the nuclear compartment, with preferential nucleolar localisation in +YB1FL and +YB1 Δ 1- Δ 6 cells (Fig. 4.18, see panels under AAV2 particle and respective merged images). This was comparable to that in transfected 293T control cells. Therefore, it was considered that rAAV2 particles and YB1 Δ 2, YB1 Δ 4, or YB1 Δ 5 were co-localised in the nucleus.

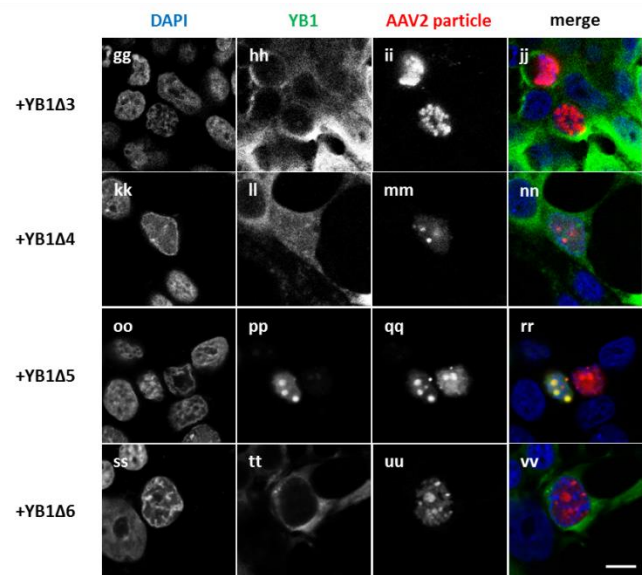
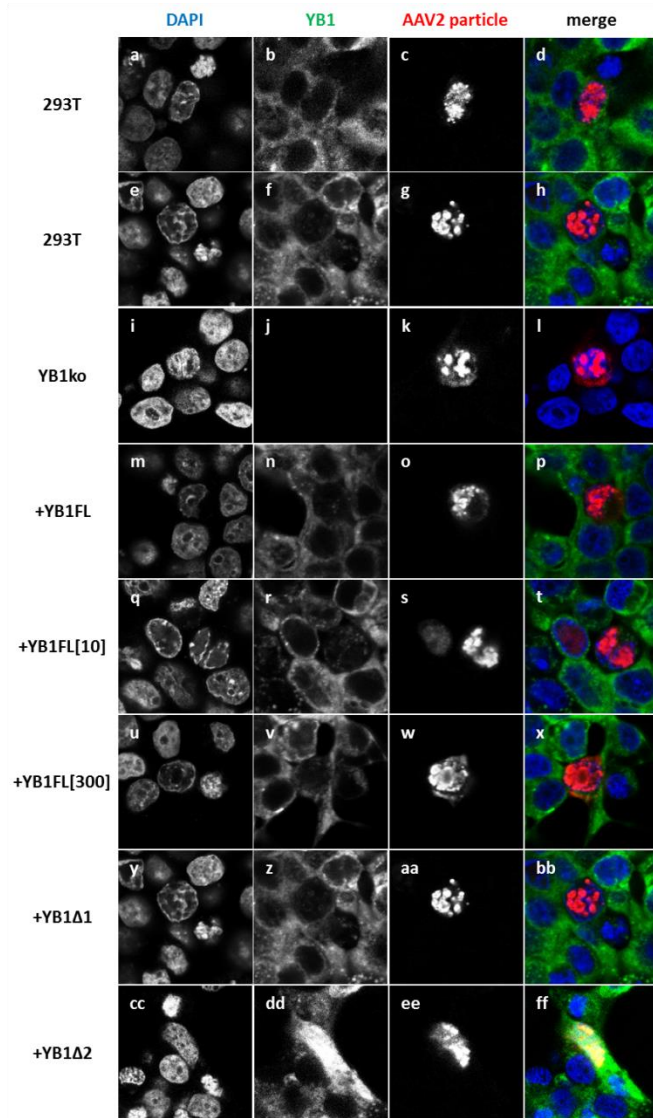


Figure 4.18 ICC and confocal laser scanning microscopy analysis of rAAV2 particle co-localisation with YB1 and YB1Δ1-Δ6 truncated mutants. Cells were transfected for rAAV2FLuc production. 72h post-transfection control 293T (**a-d** and **e-h**), YB1 knockout (**i-l**), YB1FL (**m-p**, **q-t**, and **u-x**), and YB1Δ1-Δ6 (**y-bb**, **cc-ff**, **gg-jj**, **kk-nn**, **oo-rr**, and **ss-vv**, respectively) cells were fixed, blocked and then stained for YB1 (ab114999 or PA5-19453):AAV2 particle (A20) co-localisation. DAPI-staining functioned to differentiate the nuclear compartment. Merged images of all channels are depicted. Scale bar = 10μm, and representative for all panels. Abs, antibody.

Using ICC, AAV2 Cap (Vp1-3) was detected throughout the cell, with predominant staining in the nuclear compartment, and notable exclusion from the nucleolar regions in 293T, YB1 knockout or YB1 mutants (Fig. 4.19, panels c and d). Therefore, the localisation of AAV2 Cap was considered unchanged or uninfluenced by YB1 knockout, or by the stable expression of full length YB1 or YB1 truncated mutants (Fig. 4.19, see panels under AAV2 Cap and respective merged images). Although, exceptions to this general observation included the +YB1 Δ 5 cell line and the stable expression of YB1 Δ 5, which was found to show AAV2 Cap staining within the nuclear compartment, which interestingly also included YB1 Δ 5 staining present in the nucleolar regions (Fig. 4.19, panels mm and nn). Therefore, a potential interaction and influence on AAV2 Cap localisation was inferred by the co-localisation between YB1's CTD and AAV2 Cap protein(s).

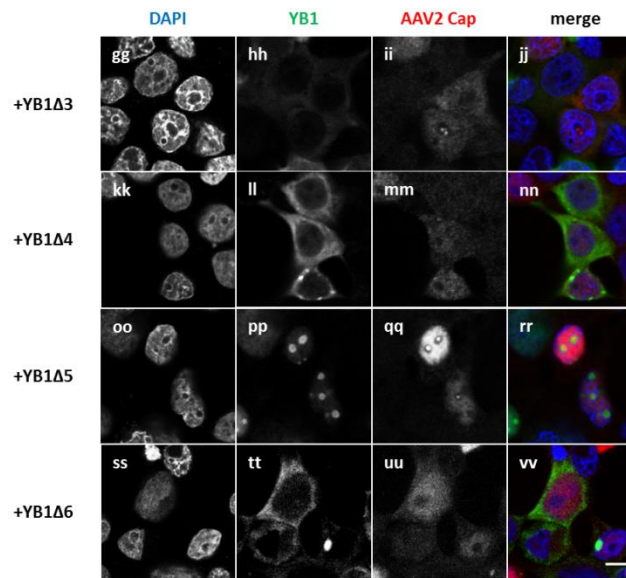
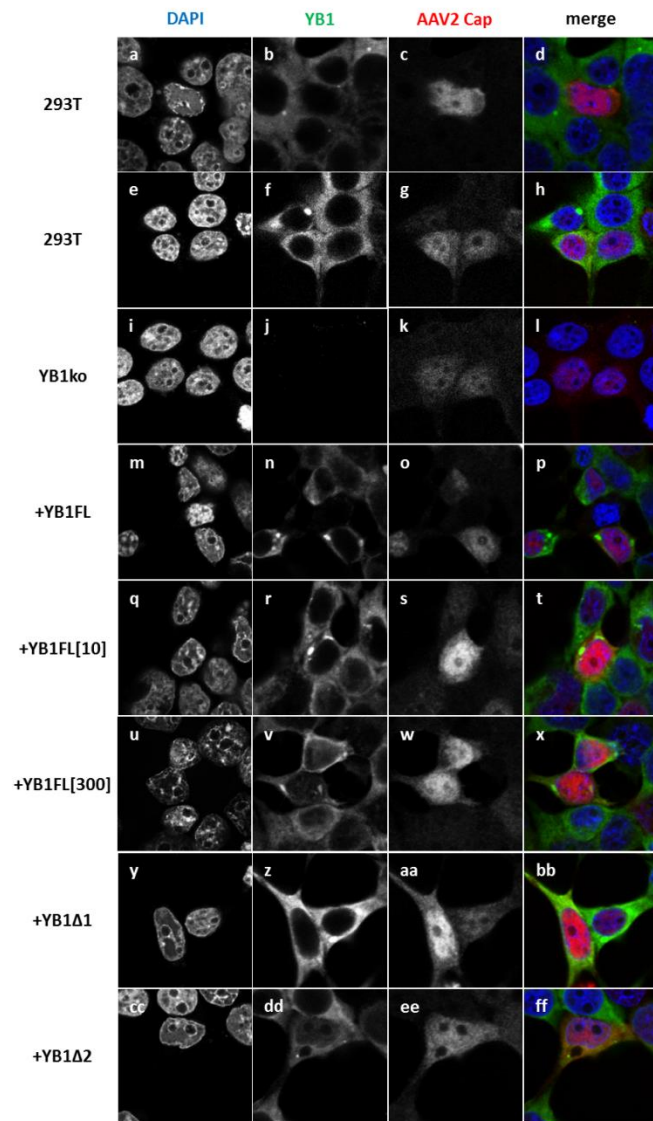


Figure 4.19 ICC and confocal laser scanning microscopy analysis of AAV2 Cap co-localisation with YB1 and YB1Δ1-Δ6. Cells were transfected for rAAV2FLuc production. 72h post-transfection control 293T (**a-d** and **e-h**), YB1 knockout (**i-l**), YB1FL (**m-p**, **q-t**, and **u-x**), and YB1Δ1-Δ6 (**y-bb**, **cc-ff**, **gg-jj**, **kk-nn**, **oo-rr**, and **ss-vv**, respectively) cells were fixed, blocked and then stained for YB1 (PA5-19453 or ab114999):AAV2 Cap (B1) co-localisation. DAPI-staining functioned to differentiate the nuclear compartment. Merged images of all channels are depicted. Scale bar = 10μm, and representative for all panels. Abs, antibody.

AAV2 Rep protein localisation was predominantly nuclear in localisation, and was relatively excluded from the nucleolar regions in transfected 293T cells (Fig. 4.20, panels c and d). Similar localisation of AAV2 Rep was observed in triple transfected YB1 knockout, +YB1FL and +YB1 Δ 1- Δ 6 cell lines (Fig. 4.20, see panels under AAV2 Rep and respective merged images). Given that YB1 Δ 2 and YB1 Δ 4 proteins were generally found in both the nucleus and cytoplasm of +YB1 Δ 2 and +YB1 Δ 4 cell lines, respectively, co-localisation between these truncated mutants and AAV2 Rep protein in the nucleus of stained cells was acknowledged.

In summary, our results suggested that YB1 has no direct influence on the localisation of rAAV2 particles, AAV2 Cap and Rep proteins during rAAV2FLuc vector production. Although, a possible influence in Cap protein localisation to the nucleolar compartments of transfected cells was acknowledged by the stable expression of YB1 Δ 5.

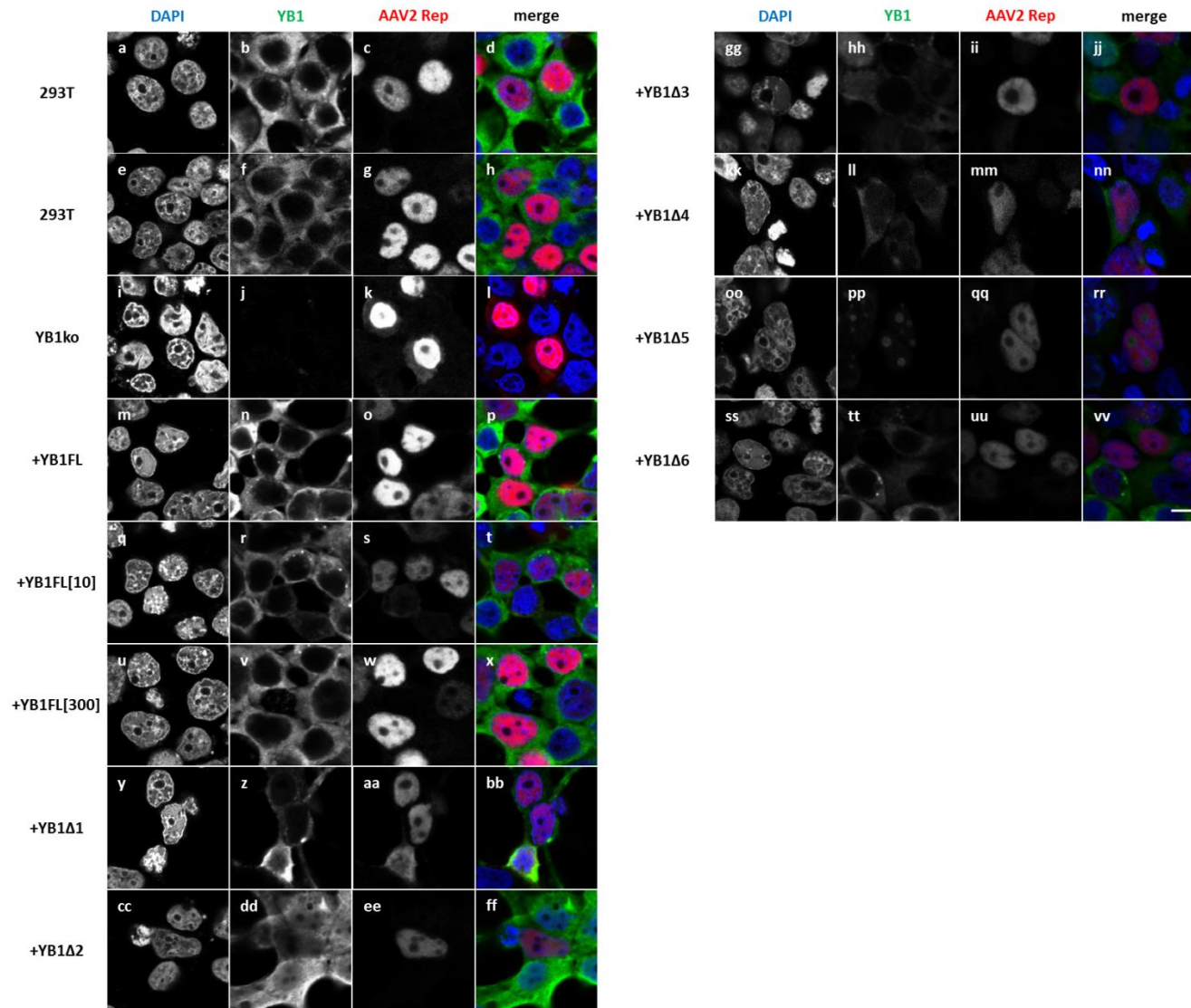


Figure 4. 20 ICC and confocal laser scanning microscopy analysis of AAV2 Rep co-localisation with YB1 and YB1Δ1-Δ6. Cells were transfected for rAAV2FLuc production. 72h post-transfection control 293T (**a-d and e-h**), YB1 knockout (**i-l**), YB1FL (**m-p, q-t, and u-x**), and YB1Δ1-Δ6 (**y-bb, cc-ff, gg-jj, kk-nn, oo-rr, and ss-vv, respectively**) cells were fixed, blocked and then stained for YB1 (PA5-19453 or ab11499):AAV2 Rep (10R-A140A) co-localisation. DAPI-staining functioned to differentiate the nuclear compartment. Merged images of all channels are depicted. Scale bar = 10μm, and representative of all panels. Abs, antibody.

YB1 is well characterised as a ss and ds DNA binding protein, with preferential binding to ssDNA (particularly the 5'-GGGGTT-3' and 5'-CCTCCT-3' motifs) and G-rich duplex sequences (Zasedateleva *et al.*, 2002; Shi *et al.*, 2013). Therefore, we next considered potential interactions of YB1 with AAV- or Adenovirus- associated DNA sequences. Table 4.2 exemplifies putative YB1 DNA binding motifs – 5'-GGGGTT-3' and 5'-CCTCCT-3', that were identified in pAAV2-MCS (*ITR* sequences), pAAV2/2-RC (*rep* sequence), and pHelper (*E2a* sequence) plasmids. Regarding the AAV2 *ITR* sequences screened in pAAV2-MCS, the putative 5'-GGGGTT-3' YB1 binding motif was identified in both *ITR* sequences of the dsDNA plasmid. This was further interpreted as present in the rAAV vector genome's left *ITR* of the +sense strand (122-127nt) and right *ITR* of the –sense strand (133-138nt from the 5'-end of the right *ITR*) as the ssDNA 5'-GGGGTT-3' binding motif. The reverse complement of the 5'-GGGGTT-3' YB1 binding motif was identified in rAAV vector genome's right *ITR* of the +sense strand, which potentiates the generation of the dsDNA 5'-GGGGTT-3' putative YB1 binding motif upon DNA replication. Examination of pAAV2/2-RC *rep* sequence revealed the putative 5'-CCTCCT-3' YB1 binding motif early in the *rep* coding sequence (2bp downstream from the ATG translation start codon). Finally, the 5'-GGGGTT-3' putative YB1 binding motif was identified 78bp upstream of the *E2a* coding sequence in the pHelper plasmid.

Table 4.2 YB1 binding motifs identified in AAV and Adenovirus sequences

target	DNA sequence
AAV2 ITR (left)	5' - GTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTA <u>GGGGTT</u> CCT - 3'
AAV2 ITR (right)*	3' - AGG <u>AACCCCT</u> AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAG - 5'
AAV2 rep	5' - ATGCC <u>GGGGTT</u> TTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCGGCATTTCTG - 3'
pHelper E2A	5' - GCGGTTAGGCTGT <u>CCTCCT</u> TCTCGACTGACTCCATGATCTTTTTCTGCCTATAGGAGAAGGAAATGGCCA - 3'

*, AAV2 right *ITR* sequence analysed in 3'>5' direction, hence reverse complement of GGGGTT binding sequence identified; bold and underlined indicates YB1 binding motif

To examine these sequences as potential YB1 binding sites, the putative YB1 binding motifs from AAV2's right *ITR*, *rep*, and Adenovirus *E2a* were PCR amplified centrally as approximately 500bp capture probes with 5'-desthiobiotin modifications (Fig. 4.21A, see left panel). Control capture probes that directly flanked the *ITR* capture probe's putative binding motif were also amplified as 500bp amplicons with 5'-desthiobiotin modification (Fig. 4.21A, see middle and right panels for left and right *ITR* flanking control capture probes, respectively). As a result, capture probes harbouring centrally located 5'-GGGGTT-3' or 5'-CCTCCT-3' putative YB1 binding motifs for *ITR*, *rep*, and *E2a* sequences demonstrated YB1 pulldown *in vitro* (Figs 4.21B-D, top panels). More specifically, the *ITR* capture probe was capable of YB1 pulldown from whole cell lysates, whereas, YB1 pulldown was not evident using YB1 knockout cell lysates from clone A2 and C5 (Fig. 4.21B, top panel). Reduced pulldown of 'YB1' was evident using B2 lysate, which was reflective of its partial YB1 knockout phenotype (Fig. Fig. 4.21B). Pulldown of YB1 from 293T lysate by the *ITR* capture probe was further substantiated due to the reduced YB1 pulldown when control capture probes (lane R+L flank) were used instead. Complete absence of YB1 pulldown when the *ITR* capture probe was omitted (lane -) was also demonstrated.

Similar pulldown of YB1 protein was evident for both the *rep* and *E2a* capture probes (Figs. 4.21C and D, respectively). Although, it was noted that the pulldown of YB1 from 293T lysate using the control capture probes was near comparable to the pulldown detected using the *rep* capture probe (see Fig. 4.21B, compare YB1 pulldown from 293T lanes using *rep* and R+L flank and capture probes). Therefore, the specificity of YB1 binding to the 5'-CCTCCT-3' motif in *rep* was suspect. Regardless, initial YB1 pulldown using PCR amplified capture probes indicated that the putative YB1 binding motifs identified in the *ITR* and *E2a* sequences may hold merit. Lysates were also verified for YB1 expression by 293T control cells or YB1 knockout, including partial knockout for clone B2 (Figs. 4.21B-D, middle panel), and normalised loads using GAPDH loading control (Fig. 4.21B-D, bottom panel). Further examination of YB1 pulldown revealed that, while YB1 pulldown was achieved using *ITR*, *rep*, and *E2a* capture probes, the left flanking control capture probe (lane L flank) was relatively unable to pulldown YB1 at the same calibre (Fig. 4.21D). This indicated that the associated YB1 pulldown by the left flanking control probe potentially represented non-specific binding. However, in contention of this, the right flanking capture probe (lane R flank) was capable of YB1 pulldown at comparable level to that of *ITR*, *rep* or *E2a* capture probes (Fig. 4.21D), despite no discernable 5'-GGGGTT-3' or 5'-CCTCCT-3'

motifs present within its sequence. The left flanking control probe, however, possessed single 5'GGGGTT3' and 5'CCTCCT3' motifs, and YB1 pulldown was significantly less by comparison.

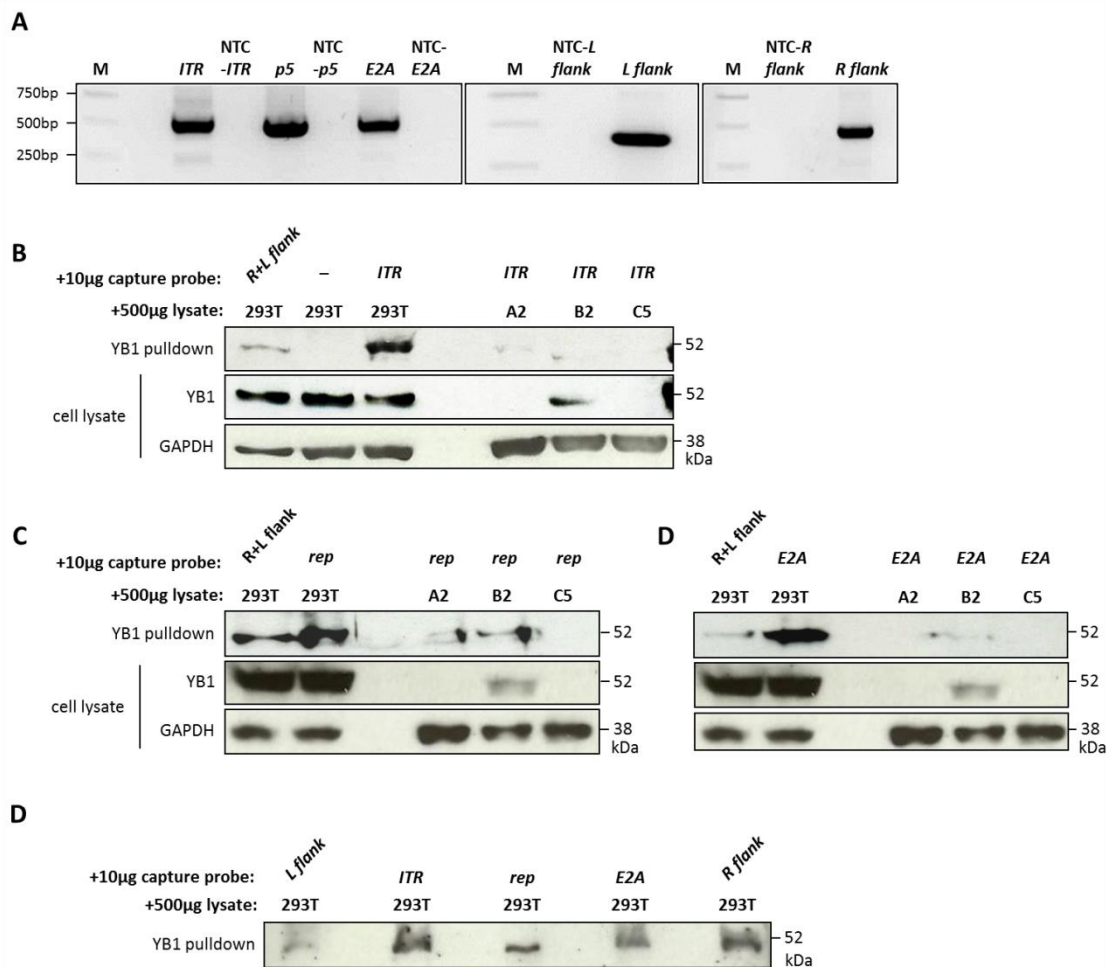


Figure 4.21 DNA affinity-YB1 pulldown demonstrates potential YB1-specific binding motifs (GGGGTT and CCTCCT) in AAV- and Adenovirus- associated DNA sequences. A) 1% TAE agarose gel electrophoresis demonstrating the expectant 5'-desthiobiotin labelled PCR amplicons corresponding to AAV2 right *ITR*, *rep*, Adenovirus *E2a* capture probes, and the immediate left *ITR* or right *ITR* flanking sequences control capture probes. PCRs were performed against non-template controls. Potential YB1 pulldown via streptavidin Dynabeads™ was identified using **(B) ITR**, **(C) p5**, or **(D) E2a** capture probes, against the pulldown of YB1 using control capture probes (mixed 1:1), and pulldown of YB1 with capture probe(s) omitted (top panel for each). Lysates were also subjected to Western blot analysis to verify YB1 expression (293T control) and YB1 knockout phenotypes for A2 and C5 clones and partial YB1 knockout phenotype in B2 clone (middle panel). Normalised protein loads were determined by GAPDH (bottom panel). **E)** YB1 pulldown from 293T lysates using *ITR*, *rep*, *E2a*, left *ITR* and right *ITR* flanking capture probes implies some non-specific binding and pulldown was achieved. L flank, left *ITR* capture probe flanking control probe; M, 1kb GeneRuler DNA Ladder; NTC, non-template control; R flank, right *ITR* capture probe flanking control probe; R+L flank, right and left *ITR* capture probe flanking control probes (1:1 mix).

The specificity of YB1 binding to putative YB1 binding motifs was further examined by using shorter 3'-biotin labelled capture probes (32bp in length). This was to focus YB1's binding to the putative YB1 binding motif found in the *ITRs*. Additional control capture probes were

introduced, including an irrelevant 32bp sequence from the *Luciferase* transgene (*Luc*), and the GC-rich sequence previously reported to bind to YB1 with high specificity (Shi *et al.*, 2013). Therefore, YB1 pulldown via *ITR*, *GC*, and *Luc* was performed in competition with cold GC-rich competitor sequence (Fig. 4.22A). It was evident that the *Luc* control capture probe was capable of YB1 pulldown in the absence of the GC-rich competitor; however, upon the addition of the GC-rich competitor, the capacity for *Luc* capture probe to pulldown YB1 was lost. This inferred that the binding to *Luc* was non-specific, and as much as twice the amount of cold GC-rich competitor to capture probe was sufficient to completely out-compete YB1-*Luc* binding. On the other hand, YB1 pulldown was evident by the GC-rich capture probe in the absence of the cold GC-rich competitor, but an equivalent level of YB1 pulldown was still observed when twice the amount of GC-rich competitor was added. Finally, the dsDNA *ITR* capture probe showed similar YB1 pulldown dynamics to the GC-rich capture probe (Fig. 4.22A). The pulldown of YB1 was evident even despite the input of GC-rich competitor at all indicated ratios. Some degree of YB1 pulldown was lost at 1:¼ and 1:½ ratio of *ITR* capture probe to GC-rich competitor, compared to the complete pulldown of YB1 at higher ratios (1:1 and 1:2). Therefore, YB1 binding and pulldown by the *ITR*'s 5'GGGGTT3' motif was not outcompeted using the cold GC-rich competitor.

The potential for YB1 binding to the ssDNA 5'-GGGGTT-3' motif found in the *ITRs* was finally examined. Firstly, the ssDNA and dsDNA (positive control) *ITR* capture probes were shown to successfully pulldown YB1 from 293T lysate in the absence of the cold GC-rich competitor (Fig. 4.22B). This binding and pulldown was, however, outcompeted at the excessive 1:10 ratio of capture probe to GC-rich competitor, and was further uninfluenced by the titration of input 293T lysate – 10-200µg (Fig. 4.22B). Next, the specificity of YB1 pulldown using ssDNA *ITR* capture probe was examined, in parallel with ssDNA *rep* and irrelevant *Luc* capture probes. Here, the ssDNA *rep* capture probe functioned as an additional control for YB1 specificity as it contains an alternative putative YB1 binding motif (5'-CCTCCT-3'). This sequence was only relevant in context with the pAAV2/2-RC plasmid (a dsDNA molecule). Therefore, YB1 pulldown was positively identified for the ssDNA *ITR* and *rep* capture probes when no GC-rich competitor was added. Interestingly, the *Luc* capture probe was incapable of YB1 pulldown, indicating that the binding to the ssDNA oligo harbouring the putative YB1 binding motifs may be highly specific. Furthermore, the degree of YB1 pulldown was more pronounced for the *ITR* capture probe compared to *rep* when no GC-rich competitor was present. When cold GC-rich competitor was added in increasing ratios to the *ITR* or *rep* capture probes (1:1, 1:5 and 1:10),

the extent of YB1 binding and pulldown was conserved with ssDNA *ITR* capture probe, especially at 1:1 ratio. However, some YB1 pulldown was recognised and detectable at 1:5 and 1:10 ratios of capture probe to GC-rich competitor. Whereas, the pulldown of YB1 via *rep* capture probe was outcompeted with the addition of the GC-rich competitor (Fig. 4.22C; 1:5 and 1:10 ratios). Therefore, given YB1 pulldown was not detectable for the irrelevant *Luc* capture probes, the binding to the 5'GGGGTT3' putative YB1 binding motif found in AAV2 *ITRs* (either ss or ds) was considered specific, inferring that YB1 may bind to the *ITR* sequences during the course of AAV processing.

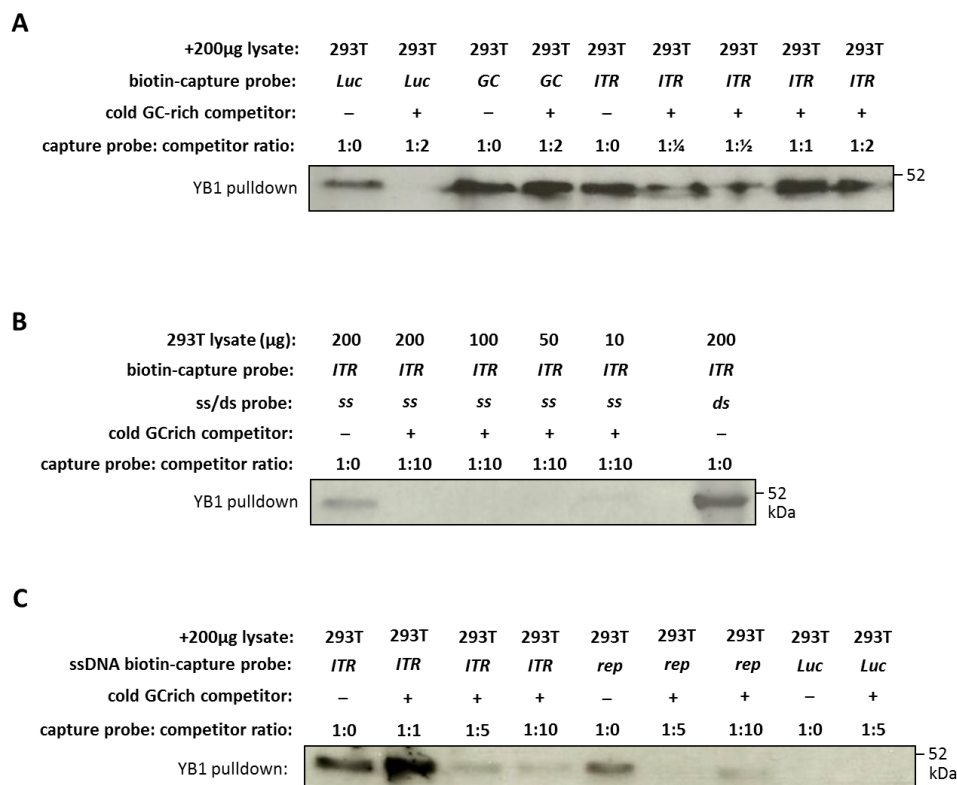


Figure 4.22 Specific binding of YB1 to putative YB1 binding motifs (GGGGTT and CCTCCT) in AAV2 *ITR* and *rep* sequences. Specific binding of YB1 to AAV2 *ITR* was determined by pulldown assays using (A) short 32nt, dsDNA 3'-biotin-labelled capture probes corresponding to the *Luciferase* control (*Luc*), GC-rich control (*GC*), and *ITR* sequences), in the presence of a GC-rich competitor at the indicated ratios. (B) ssDNA *ITR* capture probe also demonstrated the capacity for YB1 pulldown from 293T lysate), with GC-rich cold competitor. The pulldown by dsDNA *ITR* capture probe functioned as a positive control for YB1 pulldown. (C) YB1 pulldown was demonstrated by ssDNA *ITR* and *rep* capture probes, despite the addition of GC-rich competitor. The control *Luc* capture probe failed to pull down YB1.

4.2.6 Impact of YB1 knockout and rAAV2 vector production

The impact of endogenous YB1, YB1 knockout on rAAV2 protein expression was examined in triple transfected cells using PEI_{max}. Given protein loads were normalised using GAPDH as the internal loading control, the expression of Rep78, Rep52 proteins, or Vp1-3 proteins was

considered comparable between 293T and YB1 knockout cells for all harvest points over the 72h timecourse (Fig. 4.19A, top and middle panels, respectively). Non-transfected 293T and YB1 knockout cell lines as negative controls for transfection exhibited no AAV2 Rep or Vp1-3 protein expression as expected. It was also appreciated that YB1 expression was consistently detected for control 293T lysates but absent for YB1 knockout cell lysates over the 72h timecourse (Fig. 14.9A). Additionally, transfection efficiencies for pAAV2-hrGFP plasmid was measured as comparable between 293T and YB1 knockout cells (approximately 80% for both 293T and YB1 knockout) 72h post-transfection by flow cytometry (Fig. 4.19B).

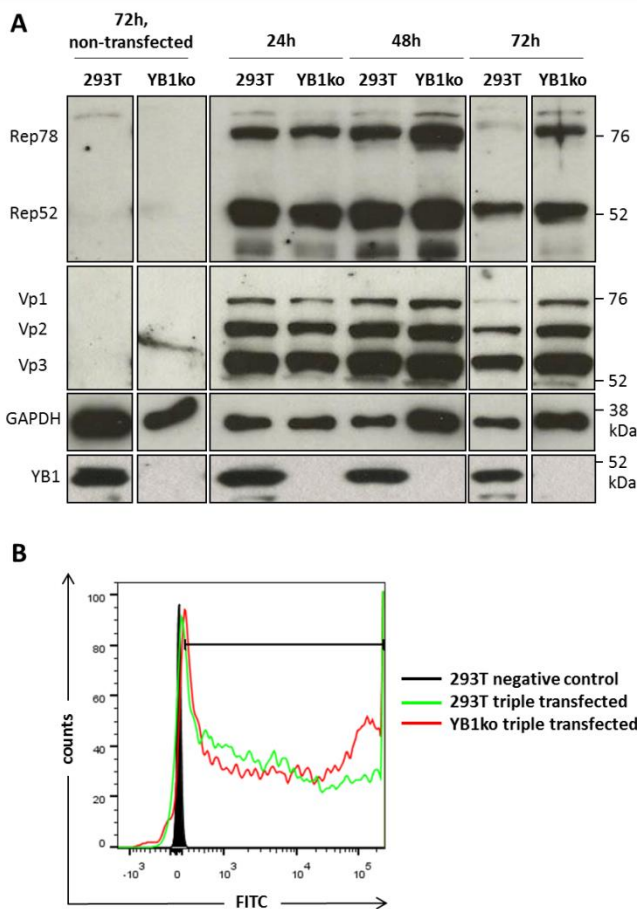


Figure 4.23 Comparable AAV2 Rep and Vp1-3 expression dynamics between 293T and YB1 knockout cells. 293T and YB1 knockout cells were triple transfected for rAAV2GFP production and lysates examined for AAV2 Rep and Vp1-3 expression. **A)** Western blot analysis revealed comparable AAV2 Rep and Vp1-3 protein expression between control 293T and YB1 knockout cells for each harvest point. This was inferred after considering GAPDH levels as an internal loading control. YB1⁺ and YB1 knockout phenotypes were also verified between control 293T and YB1 knockout cells. **B)** Flow cytometry indicated that comparable transfection efficiencies (approximately 80%) were achieved between 293T and YB1 knockout cells 72h post-transfection. Transfection efficiencies are presented in parentheses. h, hours; YB1ko, YB1 knockout.

Next, rAAV2 vectors harvested from 293T control, YB1 knockout and +YB1FL, +YB1Δ1-Δ6 cell lines were quantified using qPCR and compared to that from control 293T producer cells (Fig. 4.24). Comparable transfection efficiencies (>97%) was achieved for all cell lines tested 72h post-triple transfection for rAAV2GFP production (Fig. 4.24A). Despite this, an insignificant difference in relative rAAV2 vector genome titres was calculated for all cell lines when compared to 293T as baseline control producer cell line (n = 3 batches per cell line; *P* > 0.05). YB1 knockout cells demonstrated an average 0.85±0.19-fold difference in vector genome titres

when compared to 293T control. Similar was measured between 293T and +YB1FL, +YB1 Δ 1- Δ 6 cell lines, where average fold differences of 1.22 ± 0.15 , 0.88 ± 0.15 , 0.84 ± 0.46 , 0.69 ± 0.22 , 0.86 ± 0.3 , 0.75 ± 0.23 , 0.58 ± 0.2 , 0.73 ± 0.47 , and 1.06 ± 0.33 for +YB1FL, +YB1FL[+], +YB1FL[++], and YB1 Δ 1- Δ 5, respectively, when compared to 293T (Fig. 4.24B). Some variation in physical genome titres, for each cell line when compared to 293T control, was calculated by qPCR for each batch (Figs. 4.24C-E, for batches 1-3, respectively). Although, in general titres were calculated between a relatively narrow range of 1.2×10^7 - 6.5×10^7 vector genomes/mL across all cell lines tested. Ultimately, an insignificant and comparable effect on rAAV2 vector genome titres was displayed between all cell lines. This indicated that the YB1 knockout cell line did not correlate with an enhanced rAAV vector producer cell line, and that YB1's functional domains (recapitulated as YB1 Δ 1- Δ 6 mutants) also did not contribute significantly to rAAV2 vector processing.

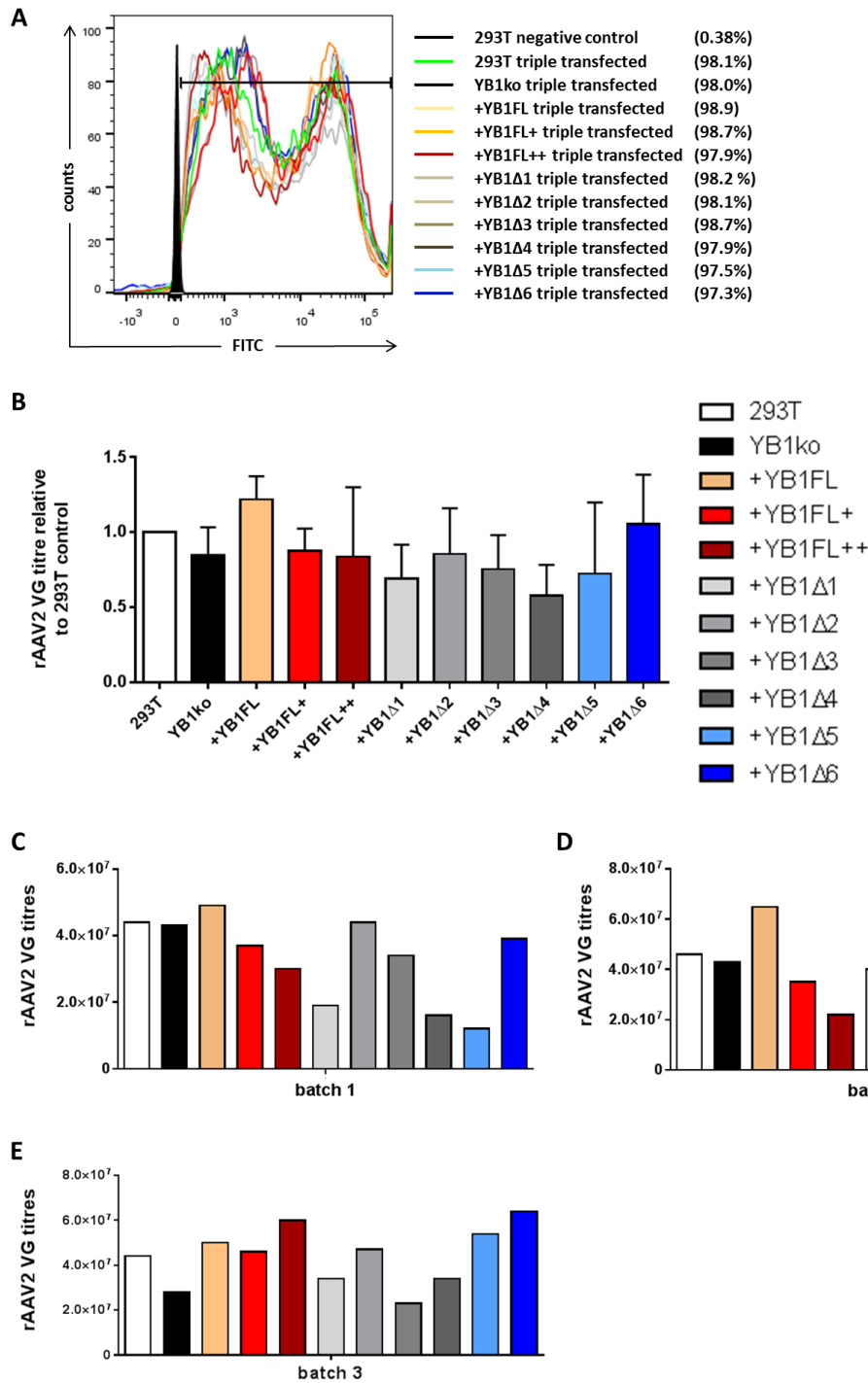


Figure 4.24 Relative rAAV2 vector genome titres from 293T control, YB1 knockout and YB1Δ1-Δ6 cell lines using qPCR. **A)** Representative flow cytometry analysis indicated that transfection efficiencies of pAAV2-hrGFP between 293T, YB1 knockout, +YB1FL and +YB1Δ1-Δ6 were comparable 72h post-transfection (>97%). This was compared to non-transfected 293T as negative control. Transfection efficiencies are displayed in parentheses. **B)** Recombinant AAV2 vector genome titres produced from YB1 knockout, +YB1FL, and +YB1Δ1-Δ6 were quantified by qPCR, and made relative to 293T as control producer cell line. A statistically insignificant difference in relative vector genome titres was calculated between YB1 knockout, +YB1FL, +YB1Δ1-Δ6 and 293T control ($P > 0.05$). Error bars reflect \pm SD from mean, $n = 3$. Individual physical rAAV2 vector genome titres/mL for each cell line is plotted for **(C)** batch 1, **(D)** batch 2, and **(E)** batch 3. VG, vector genome; YB1ko, YB1 knockout.

4.3.0 Chapter summary

With an optimal transduction system in play, the transformation of YB1 knockout cell lines with lentiviral vectors to stably express full length and truncated mutants of YB1 was unchallenged and without complications. In fact, the cohort of cells generated in the current study helped to emphasise the importance of YB1 to protect parental 293T cells from the induced cytotoxic effects of chloroquine. The rescue of the cytotoxic phenotype, as a result of chloroquine treatment of YB1 knockout cells, was mediated by the rescued expression of full length YB1. Additional resistance to chloroquine-induced cell death was mediated by YB1 mutants that commonly shared the CSD. This was particularly reflected by the YB1 Δ 2 and Δ 3 mutants, which shared a common YB1 domain – the CSD, and were the only truncated forms of YB1 that mediated resistance to chloroquine-induced cytotoxicity as opposed to rescue to wildtype or near wildtype levels like YB1 Δ 5. Ultimately, the spectrum of YB1 mutants explored suggests that the CSD was likely the principle mediator of this resistance (this was especially reflected by YB1 Δ 2- Δ 4), whereas the CTD (as reflected by YB1 Δ 5) conferred rescue from chloroquine-induced cytotoxicity to wildtype levels. Additionally, the results infer that there is perhaps a complex relationship that exists between the CSD with the A/P domain and/or CTD, which governs YB1's protective role against chloroquine-induced cytotoxicity, and is discussed further in Chapter 7 (see section 7.8 for further discussion).

The transfection of mammalian cell lines is a routine method to generate rAAV vector production. PEI-based methods are generally preferred due to its cost-effectiveness compared to lipid-based reagents. Additionally, PEI is less sensitive to pH fluctuations as exhibited by Calcium Phosphate precipitation (Park *et al.*, 2006; Hildinger *et al.*, 2007; Feng *et al.*, 2008; Ling *et al.*, 2016). Scalable generation of rAAV vector production has chiefly been based on PEI-mediated transfection of AAV plasmids, also (Grieger *et al.*, 2016). In some contrast, transfection of suspension 293 cells by Calcium Phosphate precipitation method was associated with poor transfection efficiencies and cell viability (Feng *et al.*, 2007). PEI-based transfections were performed in the current study to drive rAAV vector production between 293T and YB1 knockouts. This was due to the relatively comparable effect on cell cytotoxicity of PEI-based transfections between cell lines. Regardless, rAAV2 vector processing was considered unaffected by the YB1 knockout phenotype, with equivalent vector genome titres to parental 293T and cell lines stably expressing YB1 mutants. The expression dynamics of AAV2 Rep78 and Rep52, and Vp1-3 proteins were relatively unchanged also, between 293T and YB1 knockout cells over the course of the rAAV2 vector production. However, implied

functions of YB1 on rAAV vector processing were identified, including the colocalisation of the YB1 CTD with rAAV2 vector particles and AAV2 Cap proteins by ICC and laser scanning confocal microscopy. This was implied through colocalisation studies between rAAV2 vector particles or AAV2 Cap with YB1 Δ 5, which showed nucleolar colocalisations. Finally, we showed some specific and high affinity binding of YB1 to putative YB1 binding motifs (5'-GGGGTT-3' or 5'-CCTCCT-3') in the *ITR*, *rep* and upstream of the *E2 α* DNA sequences. Particular emphasis was stressed on the pulldown of YB1 by the *ssITR*, where YB1 could bind the D-sequence of *ITRs* and potentially interfere with (r)AAV vector processing.

Chapter 5: Identification and Molecular Cloning of *Spodoptera frugiperda* Y-Box Protein Homologue (SfYB)

5.1.0 Introduction

The production of rAAV vectors is predominantly mediated using 293T (Xiao *et al.*, 1998; Drittanti *et al.*, 2001; Liu *et al.*, 2008), Sf9 (Urabe *et al.*, 2002), or to a lesser extent Ao38 (Meghrou *et al.*, 2005) producer cells. 293T cells are typically co-transfected with plasmids harbouring the AAV2 *rep* and AAV serotype specific *cap* genes, AdV-derived helper functions, and an AAV2 *ITR*-flanking transfer vector (Xiao *et al.*, 1998). On the other hand, Sf9 cells are derived from the ovarian tissue of *S. frugiperda* (Vaughn *et al.*, 1977), while the Ao38 cell line was established from ovarian tissue of *T. ni* (Hashimoto *et al.*, 2010; Hashimoto *et al.*, 2012); both from the order *Lepidoptera*. Sf9 and Ao38 cells are well reported for the expression of large quantities of heterologous proteins using recombinant baculoviruses (Graber *et al.*, 1992; Miranda *et al.*, 1997; Usami *et al.*, 2011; Wilde *et al.*, 2014). Sf9 cells were soon after successfully exploited for rAAV vector production when co-infected with recombinant baculoviruses that include the AAV2 *ITR*-flanked transfer vector, and vector encoding the AAV2 *rep* and AAV serotype specific *cap* genes, under the control of baculovirus-derived promoters (Urabe *et al.*, 2002; Smith *et al.*, 2009).

Our current knowledge and understanding of the role host cell-derived factors towards AAV and rAAV vector biology is quite limited. This is especially the case when concerning insect cell lines such as Sf9, where very limited proteomics and genomics data are available. In fact, a large subset, if not all, of our knowledge on the impact host cellular factors has of AAV biology is modelled by human-derived cell lines (Pegoraro *et al.*, 2006; Holscher *et al.*, 2015; Mano *et al.*, 2015; Schreiber *et al.*, 2015; Satkunanathan *et al.*, 2017). Therefore, some elucidation of the nature of baculovirus infection of insect cell lines has been achieved (Joshi *et al.*, 2000; Popham *et al.*, 2010; Xue *et al.*, 2012; Dong *et al.*, 2017; Wen *et al.*, 2017; Li *et al.*, 2018), but remains far from a complete picture.

The present study aimed to identify if an Y-Box protein homologue was endogenous to insect cell lines based on *S. frugiperda* and *T. ni*, given their relevance as producer cell models for rAAV vector production (Urabe *et al.*, 2002; Meghrous *et al.*, 2005; Smith *et al.*, 2009). Particular emphasis was focused on *S. frugiperda* and the associated Sf9 cell line, which can be considered one of the most clinically relevant models for rAAV(1) vector production (Carpentier *et al.*, 2012). Here we systematically identify a set of highly conserved YB1 homologues in *Spodoptera spp.* and *T. ni in silico*, with further experimental evidence by cloning suggesting the expression of endogenous Y-Box protein homologue (SfYB) in the Sf9 cell line. We also show the expression and purification of His-tagged recombinant SfYB in Sf9 cells using recombinant baculovirus expression vector (rBEV) technology. This facilitated the generation of specific antiserum raised against the whole molecule, and in turn will further our endeavours to examine the gene regulation and/or genome editing of this factor for rAAV vector production.

5.2.0 Results

5.2.1 Prospective Y-Box or YB1 homologue(s) identified in *S. frugiperda* and *T. ni*

We sought to identify the YB1 or Y-Box protein homologue(s) in alternative cell lines (predominantly Sf9) used routinely to generate rAAV vectors (Urabe *et al.*, 2002; Meghrous *et al.*, 2005; Smith *et al.*, 2009). Firstly, Western blot analysis of Sf9 lysates (from monolayer or suspension cultures) demonstrated that a prospective signal was detected and correlated with human YB1 from 293T as positive control, using α YB1 antibodies (Fig. 5.1A). The putative YB1 homologue was resolved in a 10% SDS-PAGE gel at approximately 50kDa, which was comparable to human YB1. The α YB1 were raised against different N- and C- terminus specific epitopes of human YB1, and therefore Western blot analysis using three different α YB1 permitted greater coverage of the prospective Y-Box protein homologue in Sf9 cells, depending on the degree of homology shared to human YB1. With this in mind, the human YB1 N-terminus specific α YB1 (PA5-19453) performed consistently in detecting the prospective YB1 homologue between adherent and suspension cultured Sf9 cells (Fig. 5.1A). Whereas, human YB1 C-terminus specific α YB1 (ab114999) performed inconsistently, with signal detectable in suspension cell lysates only. This was in spite of the fact that equal loading between Sf9 lysates was achieved and determined by Coomassie[®] Blue gel staining (Fig. 5.1B). Furthermore, α YB1's (ab76149) exact epitope coverage of human YB1 was not disclosed by the manufacturers, and was difficult to interpret signals identified with this antibody further,

except that a potential Y-Box protein homologue was consistently reported in both adherent and suspension Sf9 cells.

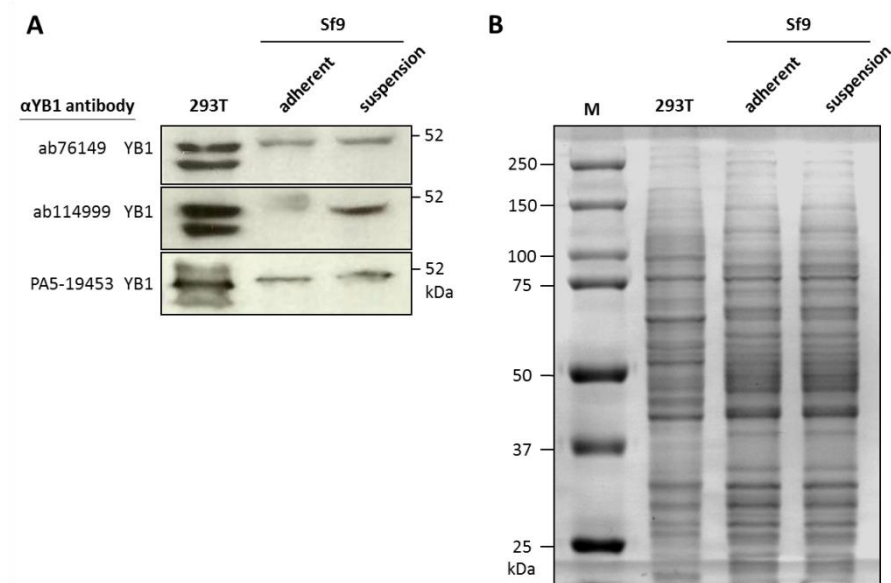


Figure 5.1 Western blot analysis identifies prospective Y-Box protein homologue in Sf9 cell lines using αYB1. A prospective YB1 homologue for *S. frugiperda* was identified in the **(A)** Sf9 cell line when cultured either as adherent monolayers or cell suspensions, that resolved to approximately 50kDa (<52kDa) by Western blotting. This was similar to human YB1 from 293T as a positive control using αYB1 antibodies. Antibodies including C-terminus specific ab76149 (undisclosed coverage), ab114999 (raised against 224-274aa of human YB1), and PA5-19453 (raised against 1-100aa of human YB1). **B)** Equal loading of protein lysates was confirmed by Coomassie® Blue gel staining, especially between adherent and suspension Sf9. M, ECL™ Rainbow™ Marker – Full Range.

With a prospective Y-Box protein homologue detected by Western blotting, we screened for the homologue sequences (transcribed RNA and aa) *in silico*. Protein BLAST results for human YB1 (NP_004550.2) resulted in no successful hits for candidate Y-Box protein homologues in *Spodoptera spp.* or *Trichoplusia spp.* Similar was reported for the additional reference control – a previously characterised YB1 orthologue (Takiya *et al.*, 2004) derived from *Bombyx mori* (silk moth), called *B. mori* Y-Box protein (BYB; NP_001036897.1). However, *in silico* BLAST analyses using the transcribed mRNA sequences (cDNA) of human YB1 (NM_004559.3) and BYB (NM_001043432.1) resulted in positive hits in the TSA database (Fig. 5.2). More specifically, input of human YB1 cDNA with organism refinement restricted to *Spodoptera spp.* resulted in positive hits for *S. frugiperda* (GESP01087352.1). A 116nt sequence match corresponded mostly to human YB1's CSD, and exhibited 97/116nt (84%) homology to *S. frugiperda*'s transcribed RNA sequence (Fig. 5.2A). Furthermore, GESP01087352.1 was recognised as an unassembled contig sequence. On the other hand, input of BYB cDNA sequence resulted in positive hits for *S. frugiperda* (GCTM01011706.1), of which 719nt

sequence match corresponded to most of BYB's cDNA, with 619/719nt (86%) homology exhibited (Fig. 5.2B). Similar was reported when organism refinement was restricted to *T. ni*, resulting in GBKU01007262.1 transcribed RNA sequence as a positive hit, with 1023nt sequence match corresponded to essentially all of BYB's cDNA, exhibiting 824/1023nt (82%) homology (Fig. 5.2C).

A

TSA: Spodoptera frugiperda TRINITY_DN35782_c0_g2_i2 transcribed RNA sequence
Sequence ID: GESP01087352.1 Length: 1565 Number of Matches: 1

Score	Expect	Identities	Gaps	Strand
110 bits(59)	2e-20	97/116(84%)	0/116(0%)	Plus/Plus
Query 153	AAGAAGGTCATCGCAACGAAGGTTTTGGAAACAGTAAATGGTTCAATGTAAGGAACGGA	212		
Sbjct 269	AAGCAGGTCATGCTGAGAAAGTTTTGGAAACCTCAATGGTTAATGTCAGAGTGGG	328		
Query 213	TATGGTTTTTCACCAAGGAATGACACCAAGGAAGATGTTATGTCACAGAGTGC	268		
Sbjct 329	TATGGTTTTTCACCAAGGAATGACACCAAGGAAGATGTTATGTCACAACTGC	384		

B

TSA: Spodoptera frugiperda SF21Tr11709 transcribed RNA sequence
Sequence ID: GCTM01011706.1 Length: 3001 Number of Matches: 3

Score	Expect	Identities	Gaps	Strand
761 bits(412)	0.0	619/719(86%)	13/719(1%)	Plus/Minus
Query 9	CGGTGAACATCGAGGAGAGGTCGTGGAGTATTTACCATCTTTGCCACCATCGCCGTTGC	68		
Sbjct 2915	CGGTGAACATCGAGGAGAGGTCGTGGAGTATTTACCATCT-AGCCACCATCGCCGTTGC	2857		
Query 69	CCAATTACACCATGGCTGATACCGAAAAGGCGCCGACGCGCAGCCcaaca--a----	122		
Sbjct 2856	CCAATTACACCATGGCTGATACCGAAAAGGCGCCGACGCGCAGCCCAACAGCAGCAGC	2797		
Query 123	aactagaaracaacacctc-a-agct----caacaagctaaacagttaaacaaagcag	176		
Sbjct 2796	AACAAGAGCAACACCCACACACACCGCAACAGCTAAAGCGCTAAACAAAAGCAG	2737		
Query 177	TCATCGCTGAGAAAGTATCGGGCACGTGAAATGGTTCAACGTCAGAGTGGATGGTT	236		
Sbjct 2736	TCATCGCTGAGAAAGTATCGGGAACCGTCAATGGTTAATGTCAAGAGTGGATGGTT	2677		
Query 237	TCATCAACAGGAATGACACCAAGGAAGATGTTTGTGCATCAGACTGCCATGCCCGTA	296		
Sbjct 2676	TCATCAACAGGAATGACACCAAGGAAGATGTTTGTGCATCAAACTGCCAATGCCCGGA	2617		
Query 297	ACAACCCACGTAAGGCTGTGCGCTCGCTCGGCAGCGAGAGGCGGTGGAGTTTGCCTGG	356		
Sbjct 2616	ACAACCCCTCGAAGGCTGTGCGCTCGCTCGGCAGCGAGAGGCGGTGGAGTTTGCCTGG	2557		
Query 357	TTGCCGGGGAGAAAGGCTTTGAAGCAGCTGGTGTACTGGTCCCAGTGGTGGAGCAGTAA	416		
Sbjct 2556	TTGCCGGGGAGAAAGGCTATGAAGCAGCTGGAGTAACTGGTCCCAGTGGTGGAGCAGTAA	2497		
Query 417	AAGGCTCACCTTATGCTGCAGACAGCGCGTGGCTACCAACCGCAATATTTCCCTCGTC	476		
Sbjct 2496	AAGGCTCGCCCTATGCTGCAGCTGACAAACCGCGCGCTCCATCGCAATATTTCCCTCGTC	2437		
Query 477	AAGGTGGCGGACGAGGCGGTGAAGGAGCTCCACGACAGAGAGGATTAGGACGTCGTGGAC	536		
Sbjct 2436	AAGGTGGCGGACGTCGGGGAGAGGCGCTCCACGACAGAGTGGAAATGGACGTCGTGGAC	2377		
Query 537	CCCCGCCAACCAAGGGGGTGCACAGAGGATGAGGGTCAAGAAAGGAGCATTGTACCAT	596		
Sbjct 2376	CCCCGCCAACCAAGGGGGTGCACAGGGGATGAGGGTCAAGAAAGGAGCAGGACCAAC	2317		
Query 597	CTCAGCGCATGTTTTCCGTCGCAATTTCCGGGTGGACCGCTGGGGAGGTCCTGGCC	656		
Sbjct 2316	CACAGCGCAGCTACTTCCGCGCAACTTCCGGGTGGACCGCTGGGGAGGTCCTGGCC	2257		
Query 657	CAATGAATAGAGGAGGGTCTCCGCGAGTGGTCCCGCAATTTCAAGCGGGGCAAGGGA	715		
Sbjct 2256	CCATGAATCGGGAGGATACCTCGCGCTCGTCCACGCACTTCAACCGGGCAAGGA	2198		

C

TSA: Trichoplusia ni UN007408 transcribed RNA sequence
Sequence ID: GBKU01007262.1 Length: 1563 Number of Matches: 1

Score	Expect	Identities	Gaps	Strand
797 bits(431)	0.0	834/1023(82%)	50/1023(4%)	Plus/Minus
Query 2	CGAAAAGCGGTGAACATCGAGGAGAGTCTGTGGAGTATTTACCATCTTTGCCACCATCG	61		
Sbjct 1469	CGAAGACCAGTGAACATCGAGGAGAGTCTGTGGAGTATTTACCATCT-AGCCACCATCG	1411		
Query 62	CCGTTCCGCAATTACACCATGGCTGATACGAAAAGGCGCCGACGCGCAGCCcaaca-	120		
Sbjct 1410	CCGTTCCGCAATTACACCATGGCTGATACGAAAAGGCGCCGACGCGCAGCCCAACAG	1351		
Query 121	-a-----caactagacacacaa-ctca-agct----caacaagctaaacagttaaacaa	169		
Sbjct 1350	CAGCAGCAACAGAGCAACAATCCCAAGCAACCGCAACAGCTAAAGCGGCTAAACAA	1291		
Query 170	aaCAGGTCATCGCTGAGAAAGTATCGGGCACTGTGAAATGGTTCAACGTCAGAGTGGG	229		
Sbjct 1290	AAGCAGGTCATCGCTGAGAAAGTATCGGGCACTGTGAAATGGTTAATGTGAAGAGTGGG	1231		
Query 230	TATGGTTTTTCACCAAGGAATGACACCAAGGAAGATGTTTTGTGCATCAGACTGCATC	289		
Sbjct 1230	TATGGTTTTTCACCAAGGAATGACACCAAGGAAGATGTTTTGTGCATCAAACTGCAATC	1171		
Query 290	GCCCGTAAACAAACCCACGTAAGGCTGTGCGCTCGTGGCGACGGGAGGCGGTGGAGTT	349		
Sbjct 1170	GCCCGAACAAACCTCGCAAGGCTGTGCGCTCGTGGCGACGGGAGGCGGTGGAGTT	1111		
Query 350	GCCGTTGGTTCGCGGGAGAAAGGCTTTGAAGCAGCTGGTGTACTGGTCCCAGTGGTAG	409		
Sbjct 1110	GCCGTTGGTTCGCGGGAGAAAGGCTTTGAAGCAGCGGAAGTGGTGTACTGGTCCCAGTGG	1051		
Query 410	CCAGTAAAGGCTCACCTATGCTGCAGCAAGCGCCGTGGCTACCAACCGCAATATTC	469		
Sbjct 1050	CCAGTAAAGGCTCGCCTATGCAAGTGCACAAACCGCCGGCTTCCATCGCAATATTC	991		
Query 470	CCTCGTCAAGGTGGCGACGAGCGGTTGAAGAGCTCCACGCAAGGAGGATTAGGACGT	529		
Sbjct 990	CCCCGTCAAGGTGGCGACGAGCGGTTGAAGAGCTCCACGTCAGAGTGGAAATGGACGT	931		
Query 530	CGTGACCCCGCCCAACCAAGGGGTTGCACAGGAGATGAGGGTCAAGAAAGGAGCATT	589		
Sbjct 930	CGTGGGCCCCCGCCCAACCAAGGGGTTGCACAGGGGATGAAGGTCAGAAAGGAGCAGGA	871		
Query 590	GTACCATCTCAGCGAGTTTTTCCGTCGCAATTTCCGCGTGGACGCGCTGGGGAGGT	649		
Sbjct 870	GCCCCACACAGCGCAGCTACTTCCGCGCAATTTCCGCGTGGACGCGCTGGTGGTGGT	811		
Query 650	CCTGCCCAATGAATAGAGGAGGGTTCCCGCAGTGGCTCGCCCAATTTCCAGGCGGGG	709		
Sbjct 810	CCGGGCTTGAATGAGGAGGGTTACCCTGCGCTGCGCAGCGCACTACCAACAGGGT	751		
Query 710	CAGGAGGACAA-AAA-T--C-AGGCT--C-AG-CGC-----CAAAATGGCAGGATGGA	754		
Sbjct 750	CAGGAGCAACCCCAACTGGCAGCAGCAACAGGCGCCGCGTCAAAACGGTCAGGAGGTA	691		
Query 755	GATGCTCAGCTACAA--CATCAA-AC--CAACAGAGGGGGCAAGCCTAA--AAGT	805		
Sbjct 690	GAGCACCACCTACCAAGTCTTCCACCAGCAACCAACAGGCAAGGCAAAAGCCCAAT	631		
Query 806	AACA-AA-CCTGTTGGTACTACTATTGAGACAACTACCAATGAGAGGCGGCTAGACA	862		
Sbjct 630	AACACCAAGCTGCTGGAGCACCATTTGAGACCACTACCAATGAGAGGCGGCTA-ACA	572		
Query 863	TCCAGAGGCGGTTATGT-GTGGACATCATTTATGTAGTGCATGTTGATTGATGTTGT	921		
Sbjct 571	CCCGTGGCTGC-ACGTTGTGGCACATC-TAAA-GTAGTGCATGTTAATCTAGTTGTGC	515		
Query 922	CACAC-TCCACCTCAACTGTCTGTCTGTACCACTGAGAGTGA-TGTTTT-T-GATG	977		
Sbjct 514	CACATATCCAACTGAAC-TGCTTGTCTGTACCACTGAAGAGTGTTCGTTTTATGTTG	456		
Query 978	TTG 980			
Sbjct 455	TTG 453			

Figure 5.2 BLAST screening identified Y-Box protein homologue cDNA sequences for *S. frugiperda* and *T. ni*. BLAST analysis was performed by input of (A) human YB1 cDNA sequence (NM_004559.3; as Query), and a small stretch of cDNA homology was identified (96/116nt; 84%) in a contig sequence derived from *S. frugiperda* (Sbjct), that corresponded to sequences that encode YB1's CSD. Whereas, input of BYB cDNA sequence (NM_001043432.1; Query) identified significant sequence homology for most of the length of BYB's cDNA for (B) *S. frugiperda* – 619/719nt (86%) homology, and (C) *T. ni* – 834/1023nt (82%) homology.

Conceptually translating GESP01087352.1 candidate cDNA for all reading frames revealed a 274aa ORF – 5'3' Frame 2 (Fig. 5.3A; see underlined sequence). Using NCBI's global align tool against human YB1 (Fig. 5.3B) or BYB (Fig. 5.3C), the protein sequences exhibited a homology of 127/334aa (38%) and 152/334aa (45%) similarity to human YB1, but 232/274aa (85%) homology or 242/274aa (88%) similarity to the BYB orthologue. Similar was obtained for GBKU01007262.1, of which prospective Y-Box protein homologue for *T. ni* was pulled out from reverse reading frame 1 (3'5' Frame 1) corresponding to a 272aa long sequence (Fig. 5.4A; see underlined sequence), and reverse BLAST against human YB1 (Fig. 5.4B) and BYB (Fig. 5.4C) presented with 124/335aa (37%) and a more significant 227/272aa (83%) homology, respectively. Therefore, probable Y-Box homologues were identified for *S. frugiperda* and *T. ni*, and thus, the encoded proteins were referred hereon as SfYB and TnYB, respectively.

A

3'5' Frame 1

DNLIALCSRYFFQWHHTNANIHRQIVIVTTKTGEHRGVEVEEFTIStopPPSPFANYTMetADTEKAPQRP
 QPQQQQQQEQSPQQAQAKAKQKQVIAEKVSGTVKWFNVKSGYGFINRNDTKEDVVFVHQTAIARN
 NPRKAVRSVGDGEAVEFAVVAGEKGFEEAAGVTGPGGEPVKGSPYAADKRRGFHRQYYPRGGGRGG
 EGAPRRGGMetGRRGPPPTQGGAGQDDEGQEGAGAPPQRSYFRNFRGGRRGGGPGPMetNRGGYRRV
 RQRNYQQGGQPQTGQPNQAPRQNGQVEAPPTSASPPQQQAKPKPNNTKPAAGTTIETTINESQA
 StopHPVAARCGTSKVVHVNSSCATYPTStopLSCLYNStopRVFRFIVVAEANNTSMetRICTYDHLDDWNVIV
 RGRLEMetELLTITDStopKCDIStopLIINYVLLMetKCFIVVPTTQCIVFFDFStopTNNFTSFGYPIStopLCNFN
 YStopNELSCRFHCPNELCRYLStopKMetIILLStopIYMetDACFNLKVFYLLStopStopIISVVKHQNSStopKLLFTC
 VKIKKKK

B

candidate
 Sequence ID: Query_233099 Length: 272 Number of Matches: 1

Range 1: 1 to 272 [Graphics](#) ▼ Next Match ▲ Previous Match

NW Score	Identities	Positives	Gaps
390	124/335(37%)	150/335(44%)	74/335(22%)
Query 1	MSSEAEQQPPAAPPALSAAADTKPGTTGSGAGSGGPGGLTSAAPAGDKKVIATKVL	60	
Sbjct 1	M A+T++ P P + A A K+VIA KV	43	
Query 61	GTWKVFNVRNGYGFINRNDTKEDVVFVHQTAIKKNIPRKYLSVGDGETVEFVVEGEKGA	120	
Sbjct 44	GTWKVFNVRNGYGFINRNDTKEDVVFVHQTAIARNIPRKAQRSVGDGEAVEFAVAGEKGF	103	
Query 121	EAAVVTGPGGVPVGSYAADRNI--HYRRYPRRRPPRNYQNSSEGEKNEGSESA	177	
Sbjct 104	EAAVVTGPGGVPVGSYAADRNI--HYRRYPRRRPPRNYQNSSEGEKNEGSESA	146	
Query 178	PEGQAQQRRPYPYRRPYRRPQYSNPPVQGEVH--EGADNQGAGEQGRPVQ	235	
Sbjct 147	P RR RR PP P QG EG + GA Q R+	184	
Query 236	NHRYGRPRFRGPPRQRQREDNEEDENQDQEQPPQRRYRNFYRRRRRPNP	295	
Sbjct 185	N+RG R GP + R ++ QG GQ P + R+N P +	241	
Query 296	KPDGKETKAADPPAENSSAPEAE-----QSGAE	324	
Sbjct 242	P ++ K P N++ P	272	

C

candidate
 Sequence ID: Query_12125 Length: 272 Number of Matches: 1

Range 1: 1 to 272 [Graphics](#) ▼ Next Match ▲ Previous M

NW Score	Identities	Positives	Gaps
1151	227/272(83%)	241/272(88%)	13/272(4%)
Query 1	HADTEKAPQPPQQQLEQQQA----QQAKTVKQKQVIAEKVSGTVKWFNVKSGYGFINR	56	
Sbjct 1	HADTEKAPQPPQQQ +Q+ Q+ QQAK KQKQVIAEKVSGTVKWFNVKSGYGFINR	60	
Query 57	NDTKEDVVFVHQTAIARNIPRKAQRSVGDGEAVEFAVAGEKGFEEAAGVTGPGGEPVKGSP	116	
Sbjct 61	NDTKEDVVFVHQTAIARNIPRKAQRSVGDGEAVEFAVAGEKGFEEAAGVTGPGGEPVKGSP	120	
Query 117	YAADKRRGYHRQYFPRQGGRRGEGAPRRGGLGRRGPPNQSGAQDGEQGGIVPSQRS	176	
Sbjct 121	YAADKRRGFHRQYFPRQGGRRGEGAPRRGGLGRRGPPNQSGAQDGEQGGIVPSQRS	180	
Query 171	FRNFRGGRRGGGPPINRGGFRVVRPRNFQAGQG---GQ-NQAQRNGQDGDASATT	231	
Sbjct 181	FRNFRGGRRGGGPPINRGGFRVVRPRNFQAGQG---GQ-NQAQRNGQDGDASATT	240	
Query 232	SN--QQGAKPKSN--KPVGTIETTINESQA	259	
Sbjct 241	++ QQQ AKPK N KP GTIETTINESQA	272	

Figure 5.4 Amino acid sequence determined for *T. ni* Y-Box homologue. *T. ni* Y-Box protein homologue transcribed RNA sequence was conceptually translated, and (A) the reverse reading frame 1 was selected as candidate for a putative Y-Box protein homologue. Reverse BLAST of the candidate aa sequence (Query) was performed using NCBI global align tool, and in turn demonstrated (B) 124/335 residues (37%) homology and 150/335 residues (44%) similarity to human YB1 (NP_004550.2; Sbjct); and (C) 227/272 residues (83%) homology and 241/272 residues (88%) similarity to BYB (NP_001036897.1; Sbjct).

5.2.2 Prospective Y-Box proteins show evolutionary conservation of their CSD

Amino acid sequence homology and conservation of Y-Box protein homologues between species was further implied through multiple-sequence alignment (MSA) that took into consideration additional Y-Box protein homologues of *Spodoptera* spp. identified *in silico*. These were identified for *S. litura*, *S. littoralis*, and *S. exigua* by methods described above, and listed in Table 5.1. As a result, alignment against the entire human YB1 and BYB calculated an overall 67% homology shared between human YB1 and selected insect homologues (Fig. 5.5A).

Considerable homology and conservation was further reported where sequences that corresponded to human YB1's CSD (residues 51-129aa) and some additional residues flanking YB1's CSD, showed conservation in Y-Box homologues for all selected sequences analysed as per heat map (Fig. 5.5A, see boxed aligned sequences: yellow-red highlighted residues were considered highly conserved). Aligned residues to YB1's CSD (residues 51-129aa) showed almost 100% conservation, especially between insect homologues, which supports the high sequence identity between YB1 or BYB and candidate SfYB or TnYB in initial BLAST assessments (Figs. 5.2, 5.3 and 5.4). Finally, the RNP-I and RNP-II binding motifs were also mapped (Fig. 5.5A, see green and blue boxed alignments, respectively) and showed almost 100% homology to human YB1's RNP-I and RNP-II binding motifs consensus sequences – K/N-G-F/Y-G-F-I/V and V-F-V-H-F (Landsman, 1992), respectively. On the other hand, conservation and homology was considered low at both termini between all Y-Box proteins (blue-green highlighted residues were considered showing no-to-low conservation). Furthermore, phylogenetic analysis of the selected MSA indicated evolutionary divergence of insect homologues from YB1 (Fig. 5.5B). Especially, the insect homologues generally clustered together under a single clade, whereas YB1 was considered an out-group, despite the significant homology observed in the CSD.

Table 5.1 List of Y-Box protein homologues explored in the present study

organism	Y-Box protein homologue	gene	cDNA (accession or GenBank)	Protein (accession)
<i>H. sapiens</i>	YB1	<i>YBX1</i> ; NC_000001.11	NM_004559.3	NP_004550.2
<i>B. mori</i> (Silkworm)	BYB	<i>BYB</i> ; NW_004582015*	NM_001043432.1	NP_001036897.1
<i>S. litura</i> (Tobacco cutworm)	not defined	<i>SLYB**</i> ; LOC111360813	XM_022977040.1	XP_022832808.1
<i>S. frugiperda</i> (Fall armyworm)	not defined	<i>SFYB**</i>	GCTM01011706.1	this study***
<i>S. exigua</i> (Beet armyworm)	not defined	<i>SEYB**</i>	GAFU01018372.1	this study***

S. littoralis

(Cotton leafworm) not defined *SLITTOYB*** EZ982856.1 this study***

T. ni

(Cabbage looper) not defined *TNYB*** GBKU01007262.1 this study***

*, whole-genome shotgun assembly derived reference sequence (unannotated); **, putatively named as no gene association in the current literature; ***, conceptually translated from cDNA (transcribed RNA sequence).

A

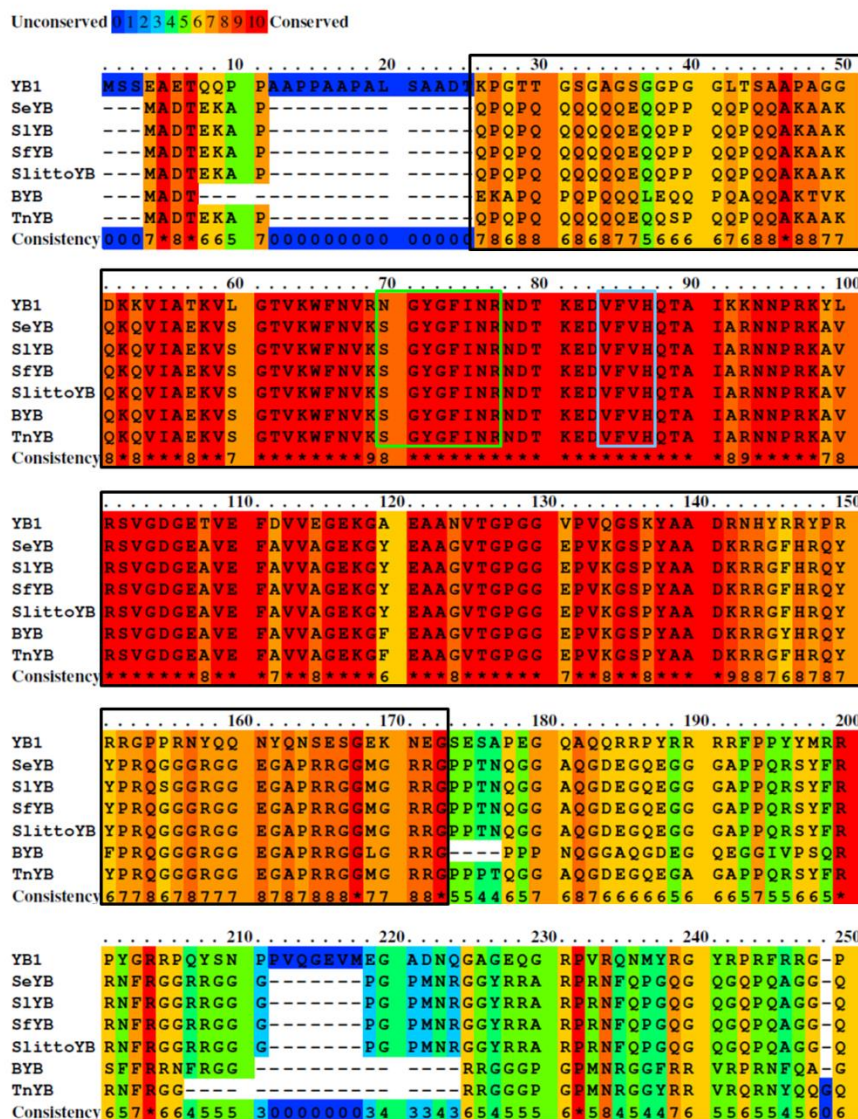
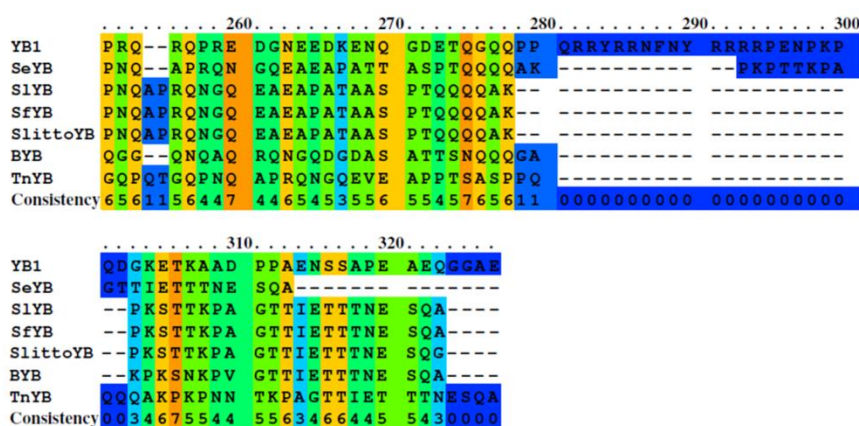


Figure 5.5 Multiple sequence alignment and phylogenetic analyses of human YB1 and Y-Box homologues from *B. mori*, *T. ni*, and *Spodoptera spp.* Figure and figure legend continues next page.

A continued



B

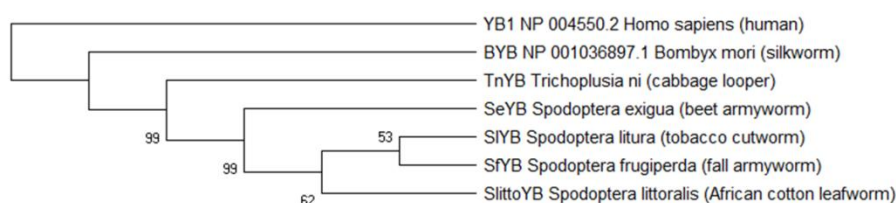


Figure 5.5 MSA of human YB1 and Y-Box homologues from *B. mori*, *T. ni*, and *Spodoptera* spp. **A)** MSA between human YB1, BYB, and prospective YB1 homologues (*S. frugiperda*, SfYB; *S. litura*, SlYB; *S. exigua*, SeYB; *S. littoralis*, SlittoYB; *T. ni*, TnYB) was performed using PRALINE MSA tool, and demonstrated considerable homology (67% across all species tested) and conservation. This was especially the case to sequences that corresponded to human YB1's CSD (51-129aa; see black boxed) and RNP-I and RNP-II binding motifs (see green and blue boxed alignments, respectively). Conservation was measured by heat-scale from 1-10 (dark blue [1] – least conserved, to crimson red [10] – complete conservation). **B)** Phylogenetic tree of selected insect Y-Box protein homologues and human YB1 aa sequences. The tree was built with MEGA X using the bootstrapped UPGMA method, and inferred evolutionary divergence of human YB1 from the insect homologues, and that the insect Y-Box protein homologues were highly conserved during the course of evolution. Numbers at nodes correspond to the % bootstrap values for 500 replicates and values <50% are collapsed.

Additionally, comparative predictive 3D-modelling of the Y-Box protein homologues was performed to reaffirm *in silico* BLAST analyses that SfYB and TnYB were Y-Box protein homologues of YB1 and/or BYB (Fig. 5.6). Firstly, YB1's N- and C- termini were previously reported as highly disordered (Guryanov *et al.*, 2012), and BYB, SfYB, and TnYB aa sequences were also identified as intrinsically disordered using the intrinsic disorder prediction algorithm based in IntFOLD (Fig. 5.6A). The N- and C- termini of each insect-derived Y-Box protein homologue were scored as highly intrinsically disordered (Fig. 5.6A; residue scores ≥ 0.5 are considered disordered). Residues 52-114 of YB1's CSD had predicted scores <0.5, and thus considered structured (Fig. 5.6A, grey shaded region), as similarly reported by Guryanov *et al.* (2012). The equivalent structured regions for BYB (residues 33-100aa), SfYB (residues 37-104aa), and TnYB (37-104) scored <0.5, also. Next, input of the entire aa sequences of YB1,

BYB, SfYB and TnYB into the PSIPRED server was performed to identify intrinsic disorder and secondary structure predictions. As a result, N- and C- termini (i.e. aa sequences that flanked the Y-Box proteins' CSD) was predicted to be disordered (Fig. 5.6B), similarly portrayed by IntFOLD (Fig. 5.6A). Human YB1 protein demonstrated no identifiable secondary structures within the N- or C- termini regions, except for β -sheets within the CSD-encoding domain (Fig. 5.6B, top panel). On the other hand, despite also displaying disordered N- and C- termini that flank a central CSD and β -sheets within the CSD, BYB, SfYB and TnYB also present with additional secondary structure predictions (Fig. 5.6B, last four panels). Specifically, BYB was predicted to harbour α -helix secondary structures at residues 21-27aa and 69-70aa; SfYB at residues 13-18aa, 26-31aa, and 73-74aa; and TnYB at residues 15-19aa, 27-31aa, 73-74aa.

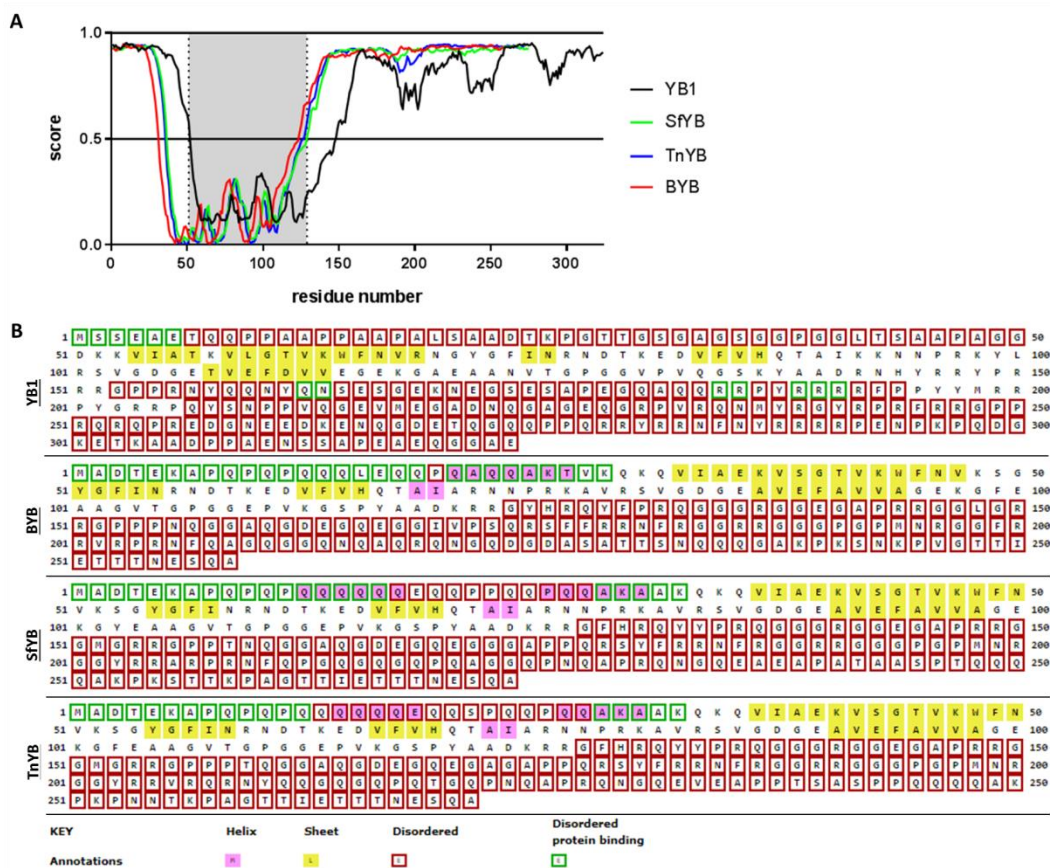


Figure 5.6 Y-Box protein homologues demonstrate conserved intrinsic disorder *in silico*. **A)** Prediction of structured and intrinsically disordered regions of human YB1 (black line), BYB (red line), SfYB (green line), and TnYB (blue line) using DISOclust server (part of IntFOLD), is depicted. Residues scoring ≥ 0.5 were considered to be disordered. YB1's CSD is highlighted by grey shading (corresponding to residues 51-129aa). Regions flanking CSDs of each Y-Box protein is considered intrinsically disordered. Intrinsic disorder was further exemplified using **(B)** PSIPRED and DISOPRED server, which also identified disordered regions in the domains that flank the central CSD (red and green boxed residues) for YB1 (top panel), BYB (second panel), SfYB (second to last panel), and TnYB (bottom panel). Figure legend continues next page.

Figure 5.6 Y-Box protein homologues demonstrate conserved intrinsic disorder *in silico*. Additionally, secondary structures were also predicted: β -sheets as per yellow highlighted residues and α -helices as per pink highlighted residues. Green boxed sequences were predicted to be disordered protein binding domains. These were chiefly found in the non-disordered CSDs. See key for annotation of secondary structures.

Predictive 3D structures of YB1, BYB, SfYB, and TnYB was then performed using the IntFOLD server, to generate template-based 3D models (McGuffin *et al.*, 2015; McGuffin *et al.*, 2018). Firstly, multi-template-based 3D modelling of YB1 and Y-Box homologues yielded 3D models with the greatest level of accuracy exhibited by the region that represented the CSDs (Fig. 5.7A-D; models are coloured by a gradient of blue [high accuracy] through white to red [low accuracy]). Therefore, the N- and C- termini were difficult to accurately predict based on template homology and conservation, which was agreeable with the intrinsic disorder exhibited by these sequences (Fig. 5.7). Furthermore, closer examination of the high accuracy modelled CSDs reflected that each Y-Box protein homologue generated similar secondary and tertiary structures to human YB1 – five β -strands were structurally predicted (Fig. 5.7A-D; see inserts and labels β 1-5). These were predicted to assemble into a highly conserved β -barrel (Fig. 5.7).

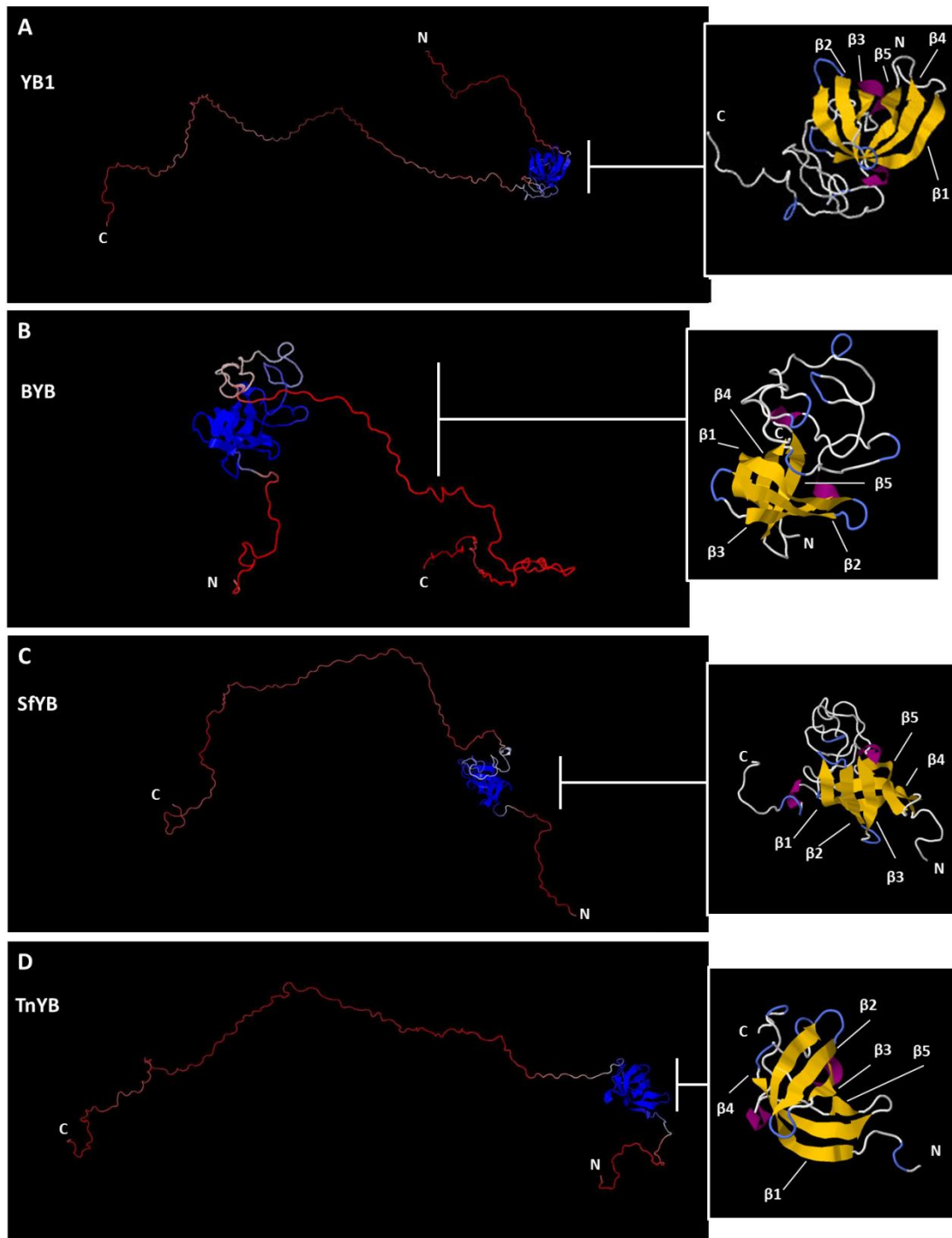


Figure 5.7 Comparative 3D-structures of YB1 and Y-Box protein homologues' demonstrate structural conservation. 3D-structures of (A) YB1, (B) BYB, (C) SfYB, and (D) TnYB Y-Box protein homologues, were modelled using the IntFOLD programme and rendered using Jmol. 3D models were confidently (global model quality score >0.34) modelled mainly for the CSD and equivalent regions in the selected homologues for predicted 3D-models, using the 5yts, 3trz, 1h95, 1mjc, 2yty, 3trs, and/or 3ulij (PDB) templates. N- and C- terminus sequences were modelled with low accuracy, as per colour gradient: blue, high accuracy; white, medium accuracy; red, low accuracy. The tertiary structure of YB1 and homologues were shown to adopt a relatively conserved β -barrel (see inserts for each). The 3D-models depicted in the inserts for each are coloured to reflect separate secondary structures: β -strands (yellow arrows; β 1-5 annotated), α -helices (pink strands), and loops (blue strands).

It was also appreciated that multi-template-based 3D modelling was able to predict tertiary structures using RNA binding and cold-shock domain containing proteins, including human YB1, as templates (see Table 5.2 for templates used to construct the top model for each Y-box protein homologue, their characteristics, and TM-scores). Chiefly, all input Y-Box protein homologues were generally templated against PDB's human YB1's CSD template (PDB template ID: 5yts; UniProt ID: P67809). Superpositions against 5yts as the top template for 3D modelling demonstrated structure conservation when concerning the length of the superimposed structures, only, i.e. when TM-align and TM-scores was normalised against the length of the template(s) used to model as opposed to the entire length of the Y-Box protein homologue that was inputted (Fig. 5.8A-D, respectively for YB1 [green model], BYB [blue], SfYB [pink], and TnYB [yellow]; see inserts of each for superimpositions against the 5yts template [green model] used to construct models, and Table 5.2, highlighted columns). This was because the templates used to model the 3D structures corresponded only to the CSDs of each Y-Box protein homologue, and N- and C- termini could not be accurately modelled. Nonetheless, use of TM-align generated high TM-scores of 0.974, 0.977, 0.921, and 0.974 for YB1, BYB, SfYB, and TnYB, respectively, along the length of the respective template (5yts) aa sequence (over 74 residues). TM-scores between $0.5 < 1.0$ are considered to be in about the same fold, with 1.0 indicated an exact match. Therefore, with high scores of $0.9 > 1.0$, BYB, SfYB and TnYB were considered to share structural conservation with YB1 (Fig. 5.7). Model qualities were calculable via the IntFOLD server, and indicated that medium-to-high model qualities were achieved (represented by global confidence score > 0.3) for the template-based models analysed. Specifically, YB1, BYB, SfYB, and TnYB generated global model quality scores of 0.344, 0.361, 0.361, and, 0.358, respectively with significant confidence in predicted 3D-structures generated for the top model constructed ($P < 0.05$ for all). Furthermore, local model qualities on a per residue basis (Fig. 5.9A-D for YB1, BYB, SfYB, and TnYB, respectively) suggests that modelling of the sequences that correspond to the CSD of each Y-Box protein homologue was modelled in best confidence. This was especially made relative to its own N- or C- disordered termini given that, on a per residue basis, the predicted residue error (the distance between two aligned molecules in Å) greatly exceeded the minimal distance (Å) exhibited by matched sequences of the Y-Box protein homologue's CSD. However, this does not come to a surprise given the IntFOLD/PDB templates only matched and modelled against the corresponding CSDs and did not include matched sequences for either termini ends.

Table 5.2 Y-Box protein model templates and TM-scores using IntFOLD and TM-align

Y-Box protein:	YB1			BYB			SfYB		
IntFOLD template number:	1	2		1	2	3	1	2	3
PDB template ID:	3trz	5yts		3trz	5yts	1mjc	3trz	5yts	1h95
protein name:	lin28a*	YB1*		lin28a*	YB1*	CspA*	lin28a*	YB1*	YB1*
host organism:	mouse	human		mouse	human	<i>Escheria coli</i>	mouse	human	human
UniProt ID:	Q8K3Y3	P67809		Q8K3Y3	P67809	P0A9X9	Q8K3Y3	P67809	P67809
TM-score ¹ :	0.371	0.227		0.47291	0.284	0.255	0.0439	0.259	0.280
TM-score ² :	0.455	0.974		0.4712	0.977	0.488	0.462	0.921	0.458

Y-Box protein:	TnYB			
IntFOLD template number:	1	2	3	4
PDB template ID:	5yts	3ulij	3ts2	2yty
protein name:	YB1*	lin28b*	lin28a*	kia0885*
host organism:	human	frog	mouse	human
UniProt ID:	P67809	B4F6I0	Q8K3Y3	O75534
TM-score ¹ :	0.270	0.307	0.332	0.210
TM-score ² :	0.974	0.455	0.329	0.339

*, CSD-containing protein; 1, TM-score using Y-Box protein homologue as normaliser; 2, TM-score using model template (reference) as normaliser.

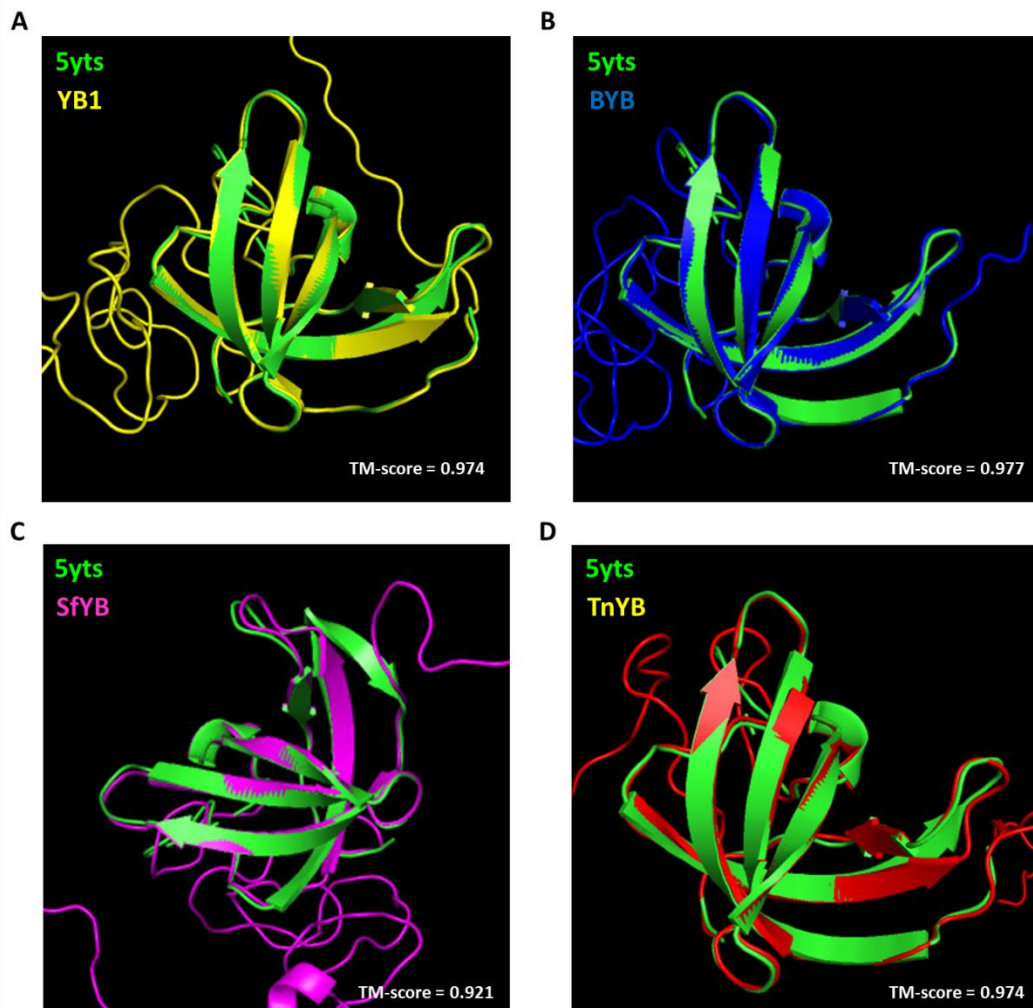


Figure 5.8 Comparative 3D-structures of YB1 and Y-Box protein homologues' CSDs demonstrate structural homology to PDB templates. Jmol rendered 3D-models of YB1, BYB, SfYB, TnYB were superimposed against the highest matching PDB template 5yts (green) from IntFOLD as reference using TM-align. Models were inputted into TM-align programme to determine CSD structure conservation for Y-Box protein homologues: **(A)** YB1 (yellow; TM-score = 0.974 over 74 residues), **(B)** BYB (blue; TM-score = 0.977 over 74 residues), **(C)** SfYB (pink; TM-score = 0.921 over 73 residues), and **(D)** TnYB (red; TM-score = 0.974 over 74 residues). TM-scores for conservation were scored between 0-1, with 1 implying a perfect match, and TM-score of $0.5 < 1$ indicating the modelled proteins are in about the same fold.

5.2.3 Recombinant SfYB(His)₁₀ expression by transient transfection of Sf9 cells

SfYB Y-Box protein homologue was confidently identified *in silico*. Given Sf9 cells are derived from *S. frugiperda* (Vaughn *et al.*, 1977), and used to generate the previously marketed, clinical grade rAAV1 vectors (Glybera) (Carpentier *et al.*, 2012) we aimed to raise antiserum specific against SfYB for further characterisation of the endogenous protein. Molecular cloning of SfYB was first achieved by codon optimising its coding sequence (*SFYBco*) for expression in *S. frugiperda*-based cell lines, with respects to codon usage and GC%. The sequence was then cloned into an expression vector designed for transfection in insect cell lines (Fig. 5.9A: a schematic overview of the cloning procedure used to construct the pEx-1.His-*SFYBco*).

Agarose gel electrophoresis demonstrated that the pExTM-1 vector backbone was successfully isolated (approximately 3834bp) from a 63bp dropout using BamHI and HindIII. Single restriction enzyme digests and uncut pExTM-1 controls did not exhibit a dropout product (Fig. 5.9B, see lanes U, +BamHI, and +HindIII, and compare to +BamHI+HindIII). Following which, the cloning and ligation of the *SFYBco* gene string fragment was confirmed by colony PCR and agarose gel electrophoresis, identifying clones 6, 7, 13, and 19 generating the expectant PCR amplicon of approximately 1111bp (Fig. 5.9C). Agarose gel electrophoresis of analytical restriction enzyme digestions further confirmed the *SFYBco* cloning in selected plasmid clones (approximately 832bp dropout expected), which was in contrast to uncut plasmid controls (Fig. 5.9D, see lanes pExTM-1 and pEx-1.His-SFYBco, under uncut, and compare to selected clones +BamHI+HindIII). Therefore, the pEx-1.His-*SFYBco* expression vector was constructed to express the *His-SFYBco* transgene, encoding rSfYB(His)₁₀.

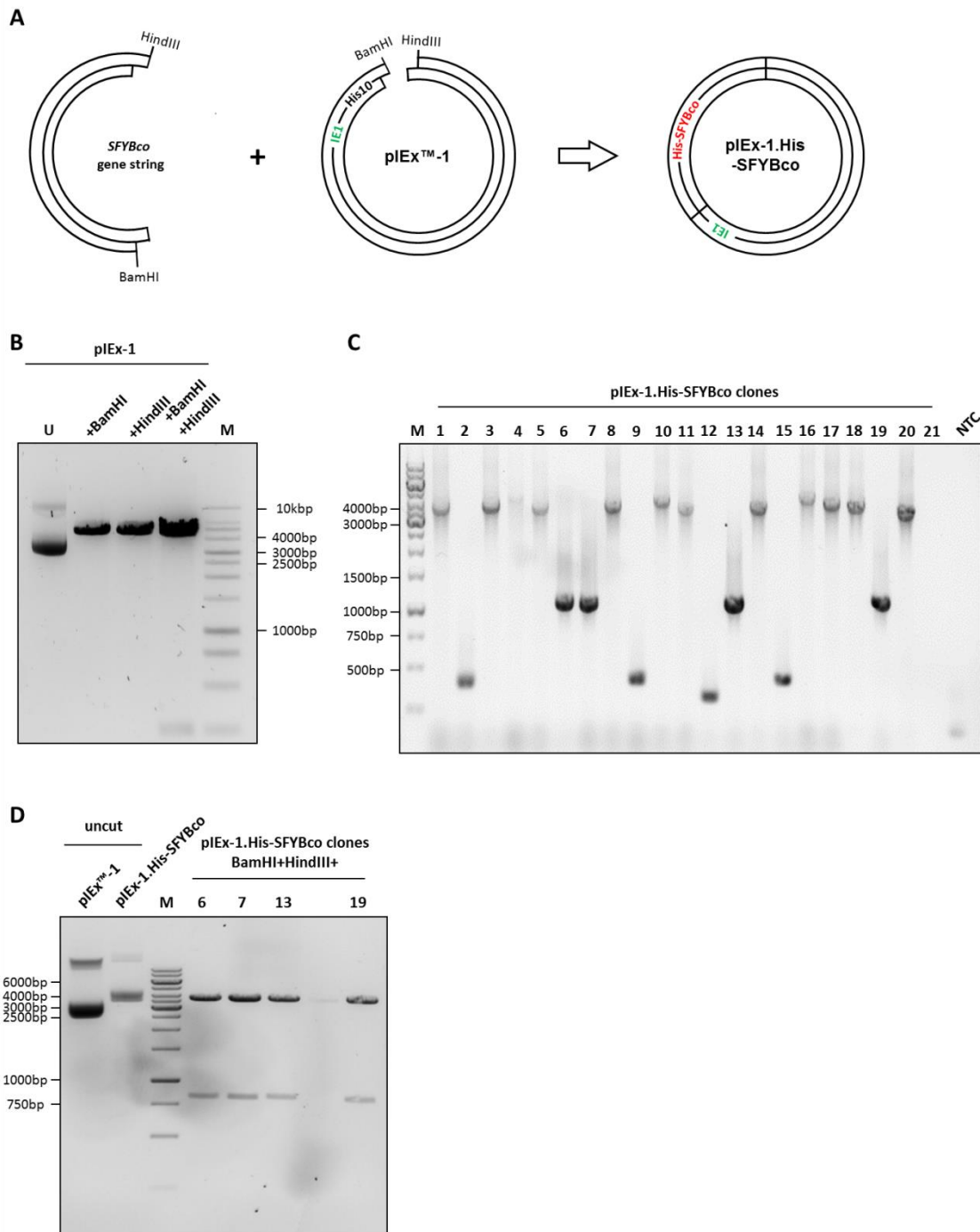


Figure 5.9 Cloning of *SFYBco* transgene into pIEx™-1 cloning vector. A) Schematic of cloning strategy used to clone *SFYBco* gene string into the pIEx™-1 cloning vector to construct pIEx-1.His-SFYBco expression vector. **B)** An approximately 62bp fragment dropout and isolation of the pIEx™-1 vector backbone was achieved by double restriction enzyme digestion, and compared to single and uncut controls by 1% TAE agarose gel electrophoresis. After transformation, positive clones 6, 7, 13, and 19 were confirmed by **(C)** colony PCR showing an expected 1111bp (approximate) PCR amplicon by 1% TAE agarose gel electrophoresis. **D)** Selected plasmid clones were further verified to produce the *SFYBco* dropout (approximately 832bp) after analytical restriction enzyme digestion, especially when compared to uncut controls. M, Generuler 1kb DNA ladder; NTC, non-template control; U, uncut.

Due to the lack of a reporter gene or selection marker in pExTM-1 plasmid, a positive control for transfection was constructed. This encoded the *EGFP* transgene under control of the IE1 promoter, which also drives *His-SFYBco* transcription in the pEx-1.His-SFYBco construct. A schematic overview for sub-cloning the *EGFP* transgene from pFBGR into pExTM-1 to generate pEx-1.eGFP is presented (Fig. 5.10A). Agarose gel electrophoresis of pExTM-1 demonstrated that the vector backbone was successfully linearized by NcoI (approximately 3897bp product expected), and was distinct from uncut control (Fig. 5.10B, see lanes U and compare to +NcoI). The pFBGR plasmid, however, harbours an additional NcoI restriction enzyme site NcoI(3) that complicated the isolation of the *EGFP* transgene as an NcoI fragment. To circumvent this, NcoI(3) was removed by AgeI and HindIII restriction enzyme digestion. The vector backbone was successfully isolated (approximately 884bp dropout expected harbouring NcoI(3)), which was in contrast to uncut and single restriction enzyme digestion controls as presented by agarose gel electrophoresis (Fig. 5.10C, see lanes U, +AgeI, and +HindIII, and compare to +AgeI+HindIII under pFBGR). The vector backbone was blunted and self-ligated, and plasmid clones of the pFBGRΔNcoI(3) intermediate vector was confirmed by analytical restriction digests. Several pFBGRΔNcoI(3) clones were encouraging as positive transformants (Fig 5.10C, clones 2, 3, 5-10), as no dropout of approximately 884bp was detected. Clone 3 of pFBGRΔNcoI(3) demonstrated no dropout of the additional NcoI digestion product of approximately 812bp in size, as would have been expected of the parental pFBGR plasmid after NcoI treatment (Fig. 5.10D, compare lanes +NcoI of pFBGR and pFBGRΔNcoI(3) clone 3). Only the dropout of approximately 922bp product corresponding to the *EGFP* transgene was observed for extraction (Fig. 5.10D, see boxed). Finally, sub-cloning of the *EGFP* transgene into the pExTM-1 vector backbone was confirmed by colony PCR with several clones (Fig 5.10E, clones 3, 5-7, and 15-17) identified as producing the expectant 745bp PCR amplicon. Remainder clones presented as concatemeric for the *EGFP* transgene (Fig. 5.10E, see lanes corresponding to clones 1, 2, 4, 8-12, 18, and 19). Therefore, successful sub-cloning of the *EGFP* transgene under promoter control of the IE1 promoter was achieved, to function as a positive control for plasmid transfections of Sf9 cells.

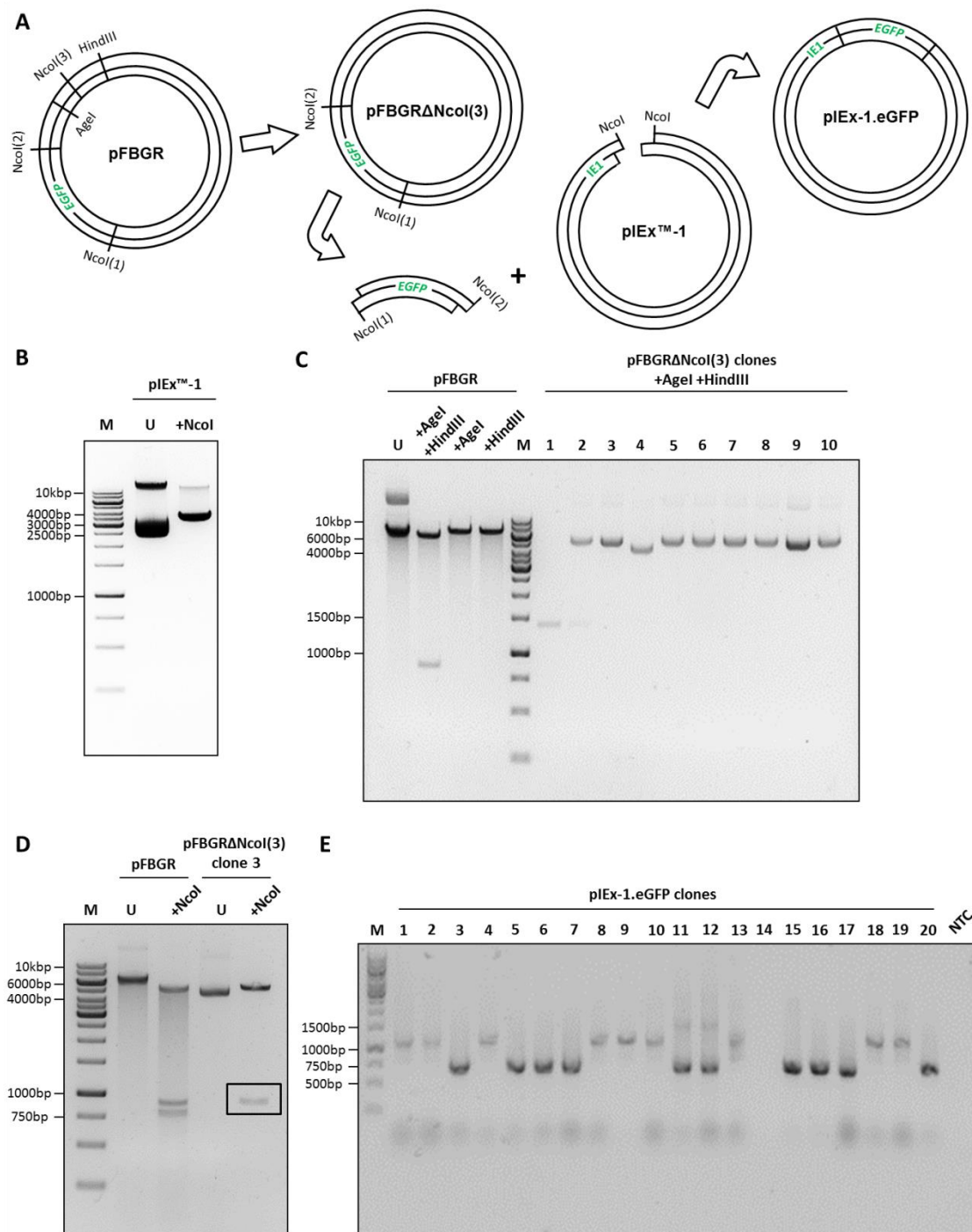


Figure 5.10 Sub-cloning of *EGFP* transgene into pIEx™-1 cloning vector. A) Schematic of cloning strategy used to clone *EGFP* transgene from pFBGR into the pIEx™-1 cloning vector to construct pIEx-1.eGFP expression vector. **B)** *Nco*I-mediated linearisation of pIEx™-1 was successfully achieved, represented by approximate 3897bp band that was distinct to the uncut control by 1% TAE agarose gel electrophoresis. **C)** The pFBGR demonstrated an approximate 884bp fragment dropout corresponding to the surplus *Nco*I(3) site using *Agel* and *Hind*III, with isolation of the pFBGR vector backbone (approximately 6355bp) compared to single and uncut controls. Additionally, pFBGRΔNcol(3) plasmid clones (2,3, 5-10) were identified by analytical restriction enzyme digestion for *Agel* and *Hind*III resistance by 1% TAE agarose gel electrophoresis. **D)** Colony PCR identified plasmid clones 3, 5-7, and 15-17 as successful pIEx-1.eGFP clones, which produced the expectant 745bp (approximate) PCR amplicon by 1% TAE agarose gel electrophoresis. M, Generuler 1kb DNA ladder; U, uncut.

Transfection of 8×10^5 monolayer Sf9 cells with pExTM-1 (mock) and pEx-1.*His-SFYBco* was performed to identify rSFYB(His)₁₀ expression. Western blotting analysis revealed that transfections with 3 μg of pEx-1.*His-SFYBco* displayed detectable signal corresponding to His-tagged protein of approximately 52kDa, 72h post-transfection (Fig. 5.11A, top panel, see boxed). The rSFYB(His)₁₀ protein was also detected using αYB1 (PA5-19453), albeit a very faint signal was produced after 3 μg plasmid transfection, especially when compared to endogenous SFYB expression (Fig. 5.11A, see black boxed for faint rSFYB(His)₁₀), and considering equal loading of lysates by Coomassie[®] Blue gel staining (Fig. 5.11B). Additionally, it was noted that non-specific signal was detected in all lysates after α6His-HRP staining, located >52kDa, and was present for non-transfected and mock transfected controls (Fig.5.11A, see NT and mock lanes, respectively).

Next, with the pEx-1.eGFP control plasmid, transfection dynamics using different quantities of pEx-1.eGFP and/or pEx-1.*His-SFYBco* plasmid (0.5, 3, and 10 μg) of 8×10^5 cells was assessed, 72h post-transfection. Flow cytometry analysis demonstrated that approximately 49.5% transfection efficiency was achieved using pEx-1.eGFP (0.5 μg), especially when compared to the 3 μg mock transfected negative control (0.36%) (Fig. 5.11C). However, low median fluorescent intensity was associated with the 49.5% transfection using pEx-1.eGFP. Additionally, input of higher quantities of pEx-1.eGFP did not positively correlate with improved transfection efficiencies of Sf9 cells – 3 μg and 10 μg of input plasmid was associated with 13.6% and 2.5% transfection efficiencies, respectively (Fig. 5.11C). His-tag staining of 0.5 μg pEx-1.*His-SFYBco*-transfected Sf9 cells exhibited 43.1% (Fig. 5.11D); whereas, His-staining was dramatically reduced despite higher quantities of input plasmid – 3 μg and 10 μg plasmid transfection demonstrated only 13.6% and 2.5% cells His-tag-positive, respectively (Fig. 5.11D, see blue and purple lines, respectively). Minimal His-staining of 15.5%, 10%, and 9.92% resulted after pExTM-1 mock transfection using 0.5, 3, or 10 μg input plasmid, respectively (Fig. 5.11D). This was all compared to 0.36%, 0.84%, and 0.65% background fluorescence for non-transfected, primary, or secondary stained-only mock transfected negative controls, respectively (Fig. 5.11D). Therefore, plasmid transfection of pEx-1.*His-SFYBco* was capable of expressing the rSFYB(His)₁₀, but at rather insufficient quantities by transient plasmid transfection using CellfectinTM II reagent at the indicated quantities.

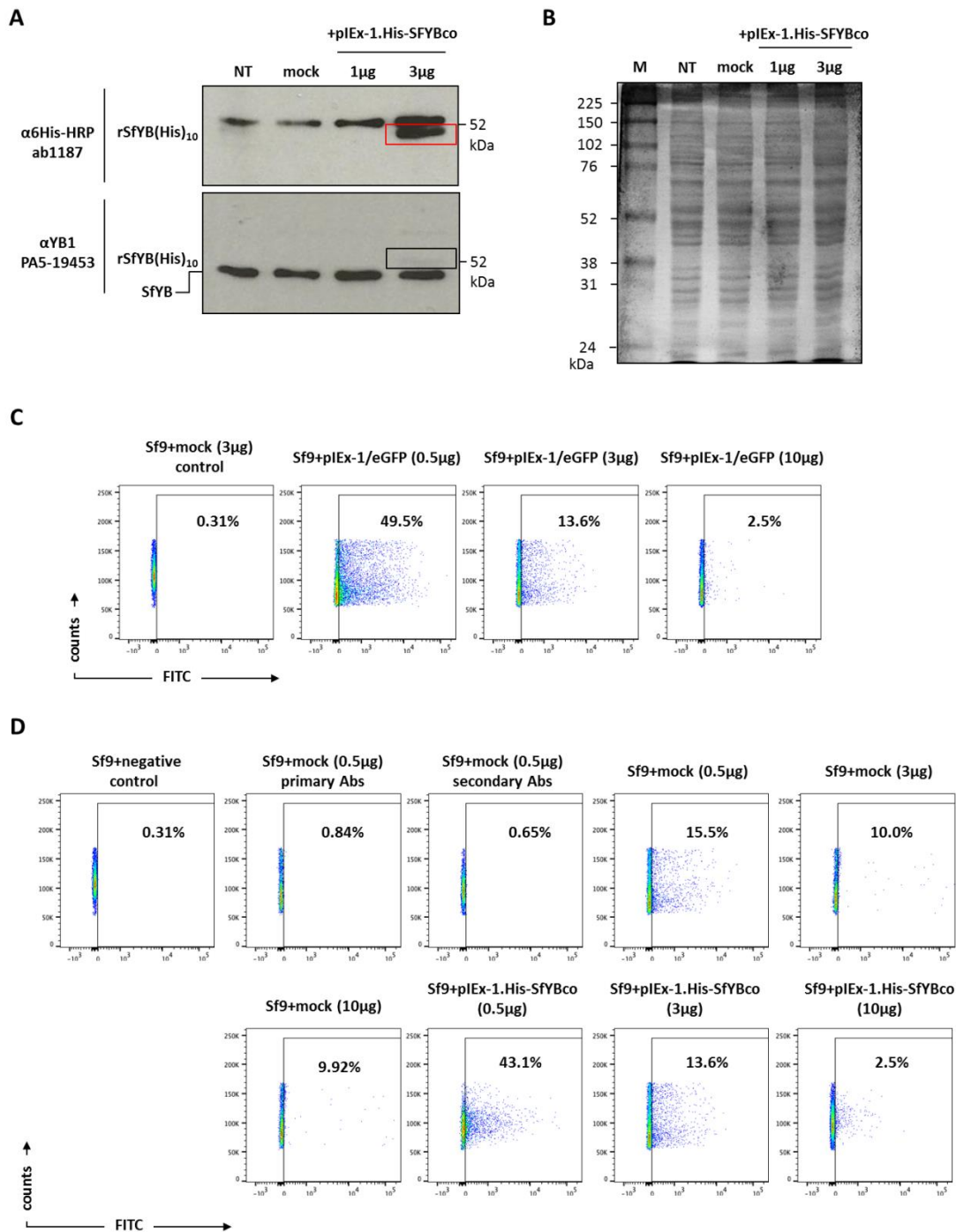


Figure 5.11 Screening for His-tagged rSfYB(His)₁₀ expression after plasmid transfection of Sf9 cells. A) Western blot analysis for rSfYB(His)₁₀ expression of non-transfected, mock transfected (pIExTM-1.1, 1 μ g), and pIEx-1.His-SfYBco transfected Sf9 cells using CellfectinTM II, harvested 72h post-transfection. Western blotting was performed using α 6His-HRP (top panel) and α YB1 (bottom panel) antibodies, and rSfYB(His)₁₀ was detected using α YB1 (boxed) as a faint signal, especially when compared to native SfYB expression. **B)** Coomassie[®] Blue gel staining demonstrated equal loading of protein samples. **C)** Flow cytometry analysis of transfection efficiencies of Sf9 cells transfected with 0.5, 3, and 10 μ g of pIEx-1.eGFP plasmid, and mock transfected (3 μ g of pIExTM-1.1). **D)** Flow cytometry analysis of non-transfected Sf9 cells and pIExTM-1.1 mock transfected cells as negative staining controls. Mock transfected and pIEx.SfYB(His)₁₀ transfected Sf9 cells (0.5, 3, and 10 μ g each) were therefore completely stained for intracellular His-tag (using α 6His-HRP) 72h post-transfection. Transfection efficiencies or His-

staining are presented as % within scatter plots. M, ECL™ Rainbow™ Marker – Full Range; NT, non-transfected.

Nonetheless, suspension Sf9 cells were transfected with pEx-1.*His-SFYBco* to express a significant amount of rSFYB(His)₁₀ for subsequent His-tag pulldown assays. Transfected Sf9 cells in suspension were first intracellularly stained, and an appreciable 48.4% and 45.3% of the transfected Sf9 cell populations were His-tag positive (Fig. 5.12A). This was especially the case when compared to non-transfected Sf9 control (0.31%) and pEx-1.*His-SFYBco* transfected cells stained with only primary (0.3%) or secondary (0.78%) antibodies as negative controls for His-tag staining (Fig. 5.12A). Following this, transfected Sf9 cells were harvested for rSFYB(His)₁₀ pulldown in duplicate, and Western blotting of 3μL and 15μL elutes for each revealed successful pulldown of His-tagged protein, which resolved to between 49-62kDa for each replicate, especially lanes corresponding to 15μL elute loads (Fig. 5.12B, see rSFYB(His)₁₀ annotation). Additional signals above and below the rSFYB(His)₁₀-annotation were also identified that may correspond to degraded rSFYB(His)₁₀, its multimerization, and/or non-specific protein pulldown (Fig. 5.12B). Furthermore, the quality of pulldown preps was assessed by Coomassie® Blue gel staining, which revealed the rSFYB(His)₁₀ protein and additional elution products only for loads corresponding to 15μL elute volumes (Fig. 5.12C, see rSFYB(His)₁₀ annotation, and red arrows – non-specific pulldown products).

Therefore, transient plasmid transfection of Sf9 cells with pEx-1.*His-SFYBco*, either as monolayer cells or in suspension, was capable of expressing the desired rSFYB(His)₁₀. Furthermore, rSFYB(His)₁₀ was determined to be a soluble protein as per pulldown assays. Transient transfection of Sf9 cells, however, only expressed modest levels of rSFYB(His)₁₀, and at impractical levels for sufficient purification for the subsequent generation of antiserum.

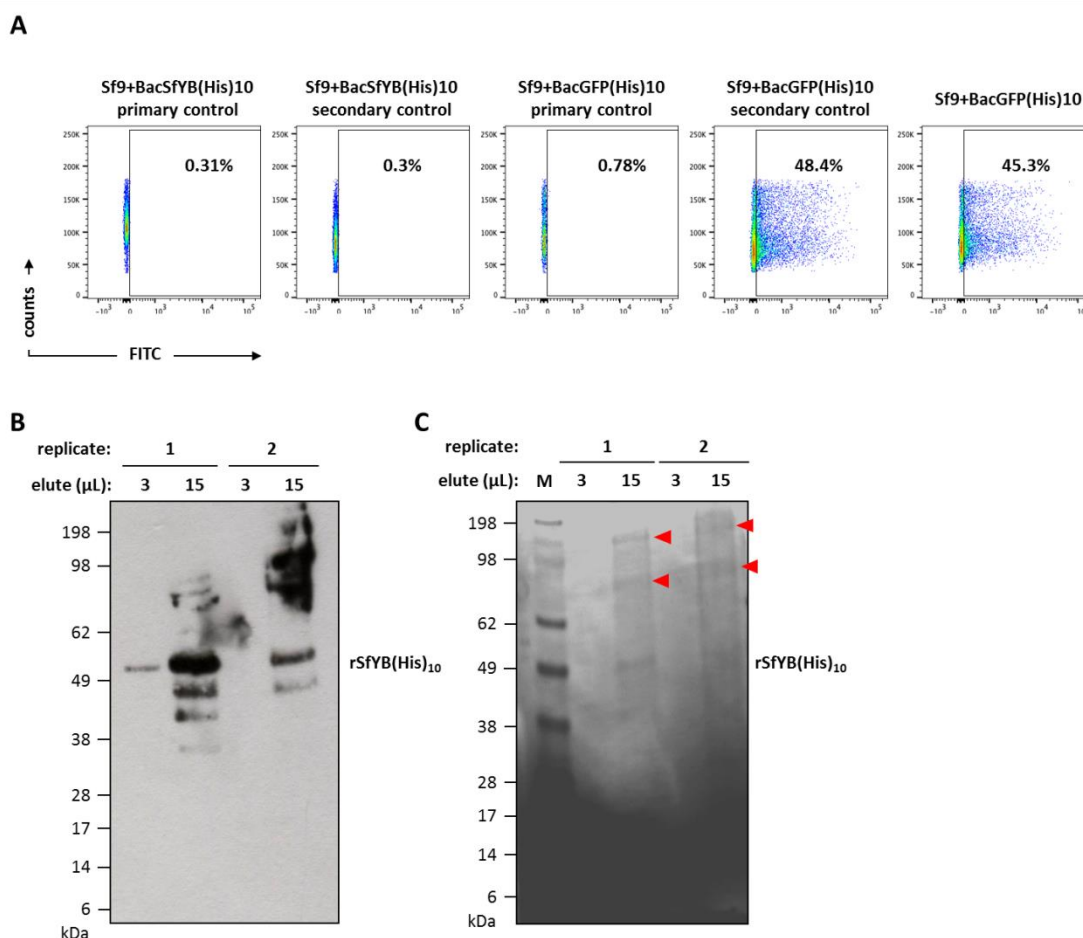


Figure 5.12 Recombinant SfYB(His)₁₀ expression and pulldown after plasmid transfection of Sf9 cells. Sf9 cells in suspension were transiently transfected with pIEx-1.His-SFYBco in duplicate, and **(A)** flow cytometry analysis demonstrated 45.3% and 48.4% of His-tag positive cells. Non-transfected and single antibody stained transfected cells served as negative controls for gating. Intracellular His-tag staining levels are presented as % of cell population of stained cells within scatter plots. Sf9 cells transfected with pIEx-1.His-SFYBco were harvested in duplicate for subsequent pulldown of His-tagged recombinant protein using Dynabeads™. Elutes of 3 and 15μL loads was subjected to **(B)** Western blot analysis which demonstrated the pulldown of rSfYB(His)₁₀ (see annotation) using α6His-HRP, and **(C)** Coomassie® Blue gel staining. This also showed that the pulldown of non-specific products (see red arrows) compromised the prep quality, but rSfYB(His)₁₀ remained distinguishable only in the 15μL elute lanes (see rSfYB(His)₁₀ annotation). Abs, antibody; M, SeeBlue™ Plus2 Pre-stained Protein Standard.

5.2.4 Scale-up production of rSfYB(His)₁₀ was achieved using recombinant baculoviruses

To achieve desirable quantities of rSfYB(His)₁₀ protein, we sought to implement rBEV technology. This involved the directional cloning by In-Fusion™ of the *His-SFYBco* transgene into a compatible baculovirus cloning vector. Therefore, with the desired homology arms incorporated, PCR amplification of the *His-SFYBco* resulted in the expected sized PCR amplicon (approximately 1044bp), confirmed by agarose gel electrophoresis (Fig. 5.13A, see inf_SFYBco lane). Subsequent cloning into a pre-linearised pTriEx™-1.1 vector backbone by In-Fusion™ resulted in the identification of four main clones 2, 5, 7, and 9 by analytical restriction enzyme

digestion (Fig. 5.13B) with approximately 458bp fragment expected. This was in contrast to double or single restriction enzyme digested controls of pTriEx™-1.1 (Fig. 5.13B, see lanes +EcoRI+NcoI under pTriEx-1.His-SFYBco clones 2, 5,7, and 9, and compare to +EcoRI, +NcoI, and +EcoRI+NcoI lanes under pTriEx™-1.1) and single restriction enzyme digested controls of the selected pTriEx-1.His-SFYBco clone 2 (Fig. 5.13B, see lanes +EcoRI+NcoI under pTriEx-1.His-SFYBco clones 2, 5, 7, and 9, and compare to +EcoRI or +NcoI lanes under pTriEx-1.His-SFYBco clone 2). Additionally, directional cloning by In-Fusion™ was confirmed by Sanger sequencing analysis for clones 2, 5, 7, and 9 (Figs. 5.13C-F, respectively). This revealed that the 5'-cloning junction for all selected clones of the pTriEx-1.His-SFYBco plasmid was in-frame to its start ATG codon (Figs. 5.13C-F, see start codon annotation in relation to *His-SFYBco* coding sequence, underlined green).

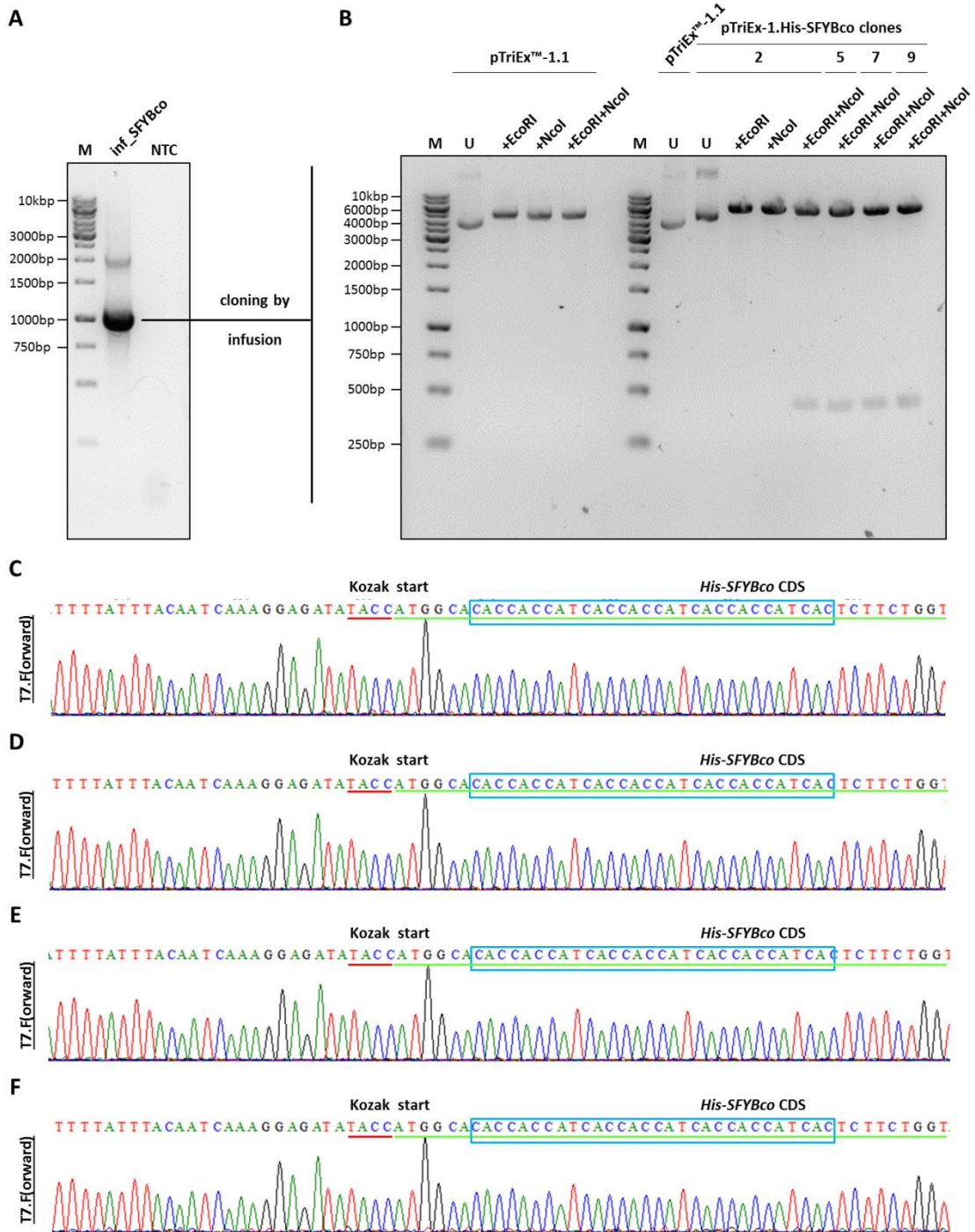


Figure 5.13 Cloning by In-Fusion™ of *His-SFYBco* transgene into pTriEx™-1.1 baculovirus cloning vector. **A)** *His-SFYBco* transgene sequence with the required homology arms was amplified and confirmed by 1% TAE agarose gel electrophoresis (inf_SFYBco; approximately 1044bp amplicon expected), relative to non-template control. **B)** Selected clones of pTriEx-1.His-SFYBco plasmid were screened for approximately 458bp fragment dropout after EcoRI and NcoI restriction enzyme digestion, and compared to uncut and single cut controls. This indicated all selected plasmid clones were cloned correctly. Plasmids were further verified by (C-F) Sanger sequencing using T7.F(oward). The 5' cloning junction was confirmed to be in-frame of the His-tag and SfYB encoded sequences for pTriEx-1.His-SFYBco plasmid clones **(C)** 2, **(D)** 5, **(E)** 7, and **(F)** 9. Kozak sequence (underlined red), start ATG codon,

10x His-tag coding sequence (blue boxed), and *His-SFYBco* CDS (underlined green), are annotated. CDS, coding sequence; M, Generuler 1kb DNA ladder; NTC, non-template control; U, uncut.

Recombinant BEVs encoding the *His-SFYBco* transgene and the control rBEV encoding *His-GFP* were then produced by co-transfection of Sf9 cells with *flashBAC*[™] DNA and baculovirus cloning vectors. Although accelerated P1 rBEV stocks were harvested from the culture media, Sf9 P1 producer cells for each BacSfYB(His)₁₀ vector clone (2, 5, 7, and 9) were harvested after 7 days post-infection. Both rSfYB(His)₁₀ and rGFP(His)₁₀ proteins were detected using α 6His-HRP by Western blotting (Fig. 5.14A, top panel); rSfYB(His)₁₀ was identified to resolve to approximately 50kDa as previously observed using transient transfection methods. This signal was absent in non-infected and BacGFP(His)₁₀-infected Sf9 controls. Additionally, the rGFP(His)₁₀ protein was identified between 25-37kDa, which was in agreement with previous reports concerning poly-His-tagged GFP proteins with approximate molecular weight of 30kDa (Mohamadipoor *et al.*, 2009). However, it was noted that non-specific signals were detected as previously observed using transient transfection of Sf9 cells (Fig. 5.11A, non-specific signal was identified >52kDa) even in non-infected control Sf9 cells. An additional non-specific signal was picked up in baculovirus-infected cells at approximately 37kDa, but absent from non-infected Sf9 control. These did not affect our analysis that the rSfYB(His)₁₀ and rGFP(His)₁₀ was detectable after infection with rBEVS, because the rSfYB(His)₁₀ protein expression was identified in BacSfYB(His)₁₀-infected Sf9 P1 producer cells only by Western blotting using α YB1, and thus was not expressed in non-infected Sf9 and BacGFP(His)₁₀-infected Sf9 control cells (Fig. 5.14A, bottom panel). Staining with α YB1 also permitted the detection of the endogenous SfYB expression in all control and infected Sf9 cells, which resolved below the rSfYB(His)₁₀ protein – approximately 50kDa (Fig. 5.14A, bottom panel). Especially, equal loading between samples was demonstrated by Coomassie[®] Blue gel staining (Fig. 5.14B), and infection efficiency using BacGFP(His)₁₀ was considered sufficient at 88.2% by flow cytometry analysis (Fig. 5.14C).

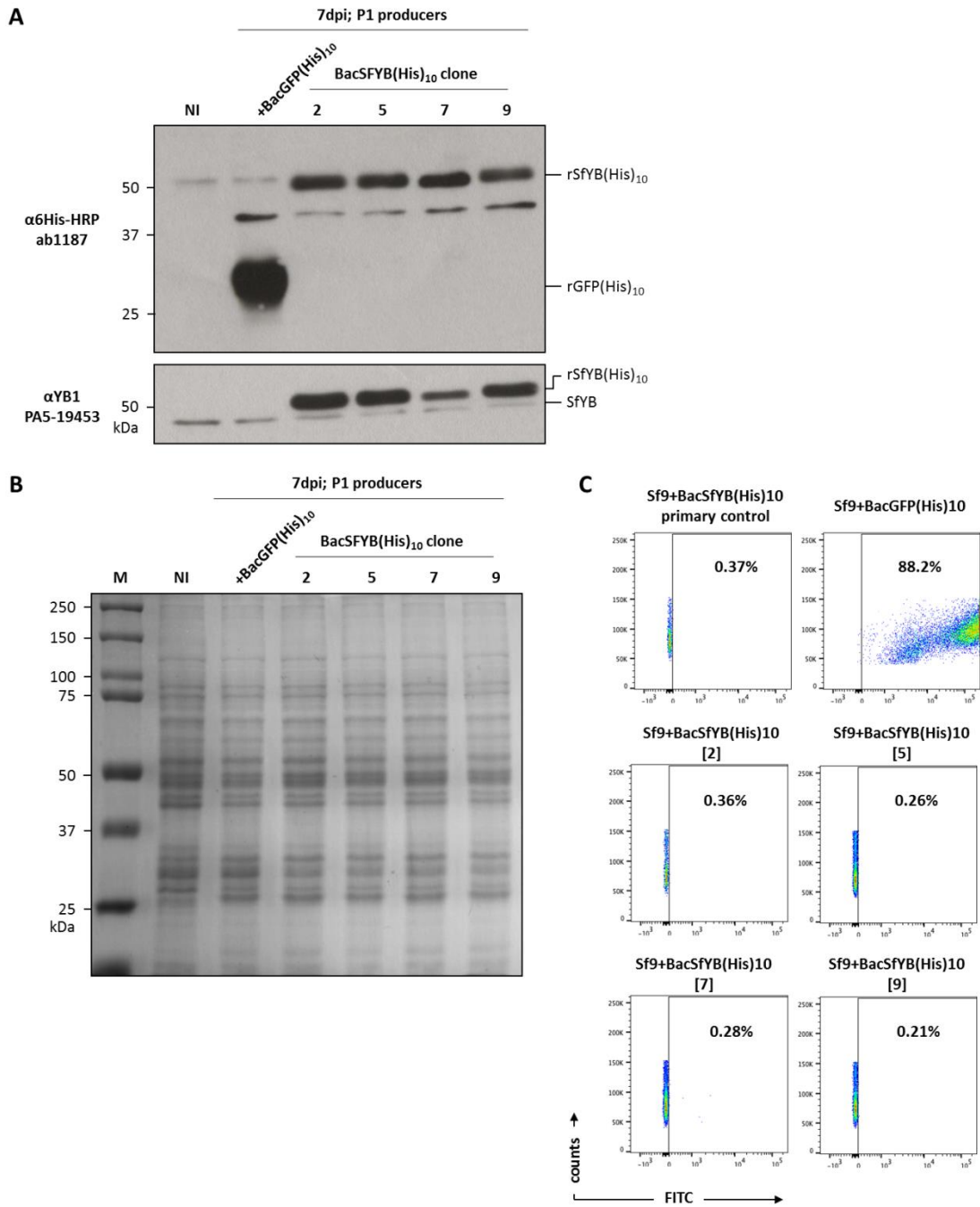


Figure 5.14 Recombinant SfYB(His)₁₀ expression confirmed after BacSFYB(His)₁₀ infections. Sf9 cells infected with P0 BacSFYB1(His)₁₀ (P1 producers) were harvested and subjected to **(A)** Western blotting using α 6His-HRP and α YB1 (top and bottom panels respectively). The expression of rSFYB(His)₁₀ (see annotation) from each BacSFYB(His)₁₀ vector clone-infection of Sf9 cells, and rGFP(His)₁₀ control (see annotation), was evident. Equal loading of protein samples was confirmed by **(B)** Coomassie® Blue gel staining. **(C)** Flow cytometry for rBEV-infected Sf9 cells revealed infection efficiency of 88.2% with BacGFP(His)₁₀ positive control, and compared to Sf9 infected with BacSFYB(His)₁₀ vector clones 2, 5, 7, and 9 (ranging between 0.21-0.36%) and non-infected Sf9 cells served as a negative control for gating. Infection efficiencies are presented as % within scatter plots. 7dpi, 7 days post-infection; NI, non-infected; M, Precision Plus Protein™ Dual Colour Standards.

Sf9 cells were further assessed by intracellular His-tag staining and flow cytometry 72h post-infection with BacGFP(His)₁₀ and selected BacSfYB(His)₁₀ vector clones. Sf9 infected with BacSfYB(His)₁₀ vector clone 2 or BacGFP(His)₁₀ stained with either primary or secondary antibodies functioned as staining controls, and demonstrated negligible FITC- or APC- positive populations (Figs. 5.15A and B, respectively; see black and grey lines for both histograms). The single staining controls for BacGFP(His)₁₀-infected Sf9 cells reported fluorescence representing infection efficiency of approximately 82% for BacGFP(His)₁₀, similarly to unstained BacGFP(His)₁₀-infected Sf9 (Fig. 5.15A, see light grey, dark grey and green lines, respectively). Intracellular staining of His-tagged recombinant protein was therefore reported as between 3.86%-29.6% for BacSfYB(His)₁₀-infected Sf9 cells despite the representative 82% infection efficiency (Fig. 5.15A, see red, orange, dark blue and light blue lines, for Sf9 infected with BacSfYB(His)₁₀ clones 2, 5, 7, and 9, respectively). BEV clone 9 infections presented with the least intracellular His-tag staining (3.86%), and BEV clone 7 presented as the best performer (29.6%). Although, interestingly, use of Alexa Fluor-647 secondary for His-tag staining presented rather contradictory staining levels - GFP(His)₁₀ was reported at 40.3% (Fig. 5.15B, green line), despite 83.2% infection efficiency (Fig. 5.15A, green line). Additionally, the degree of His-tag positive cells in BacSfYB(His)₁₀ was dramatically reduced to only 1-4.5% of the stained cell populations (Fig. 5.15B, see red, orange, dark blue and light blue lines, for Sf9 infected with BacSfYB(His)₁₀ clones 2, 5, 7, and 9, respectively). Therefore, intracellular staining for rSfYB(His)₁₀ via the His-tag was insufficient or unreliable in correlating expression levels. This was perhaps a result of disorder and unpredictable folding that occurs at either N- or C-terminus of SfYB protein (Fig. 5.6). However, the BacSfYB(His)₁₀ vector clone 7 infections, at the very least, out-performed the remainder and was used in further infection studies.

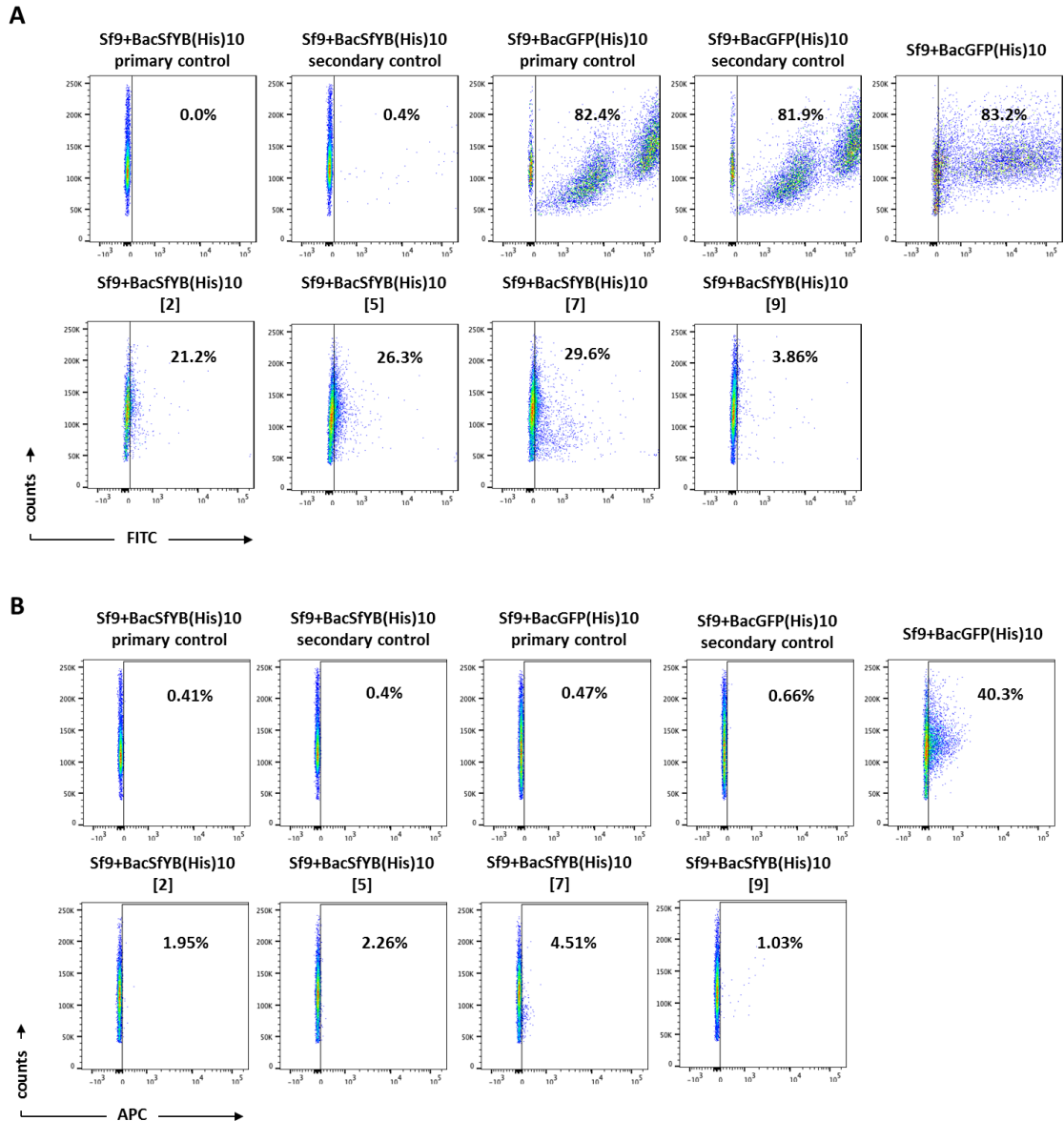


Figure 5.15 Intracellular staining for His-tag protein may not correlate expression levels for target rSfYB(His)₁₀. Sf9 cells were infected with BacSfYB(His)₁₀ and BacGFP(His)₁₀ and His-tagged recombinant protein expression was analysed by intracellular staining using α 6His-HRP primary and **(A)** Alexa Fluor 488-conjugated secondary. Intracellular His-staining was deduced to range between 3.86-29.6% for BacSfYB(His)₁₀ infected cells, despite a representative 83.2% infection efficiency with BacGFP(His)₁₀. Otherwise, cells were stained with **(B)** Alexa Fluor 647-conjugated secondary, and revealed low 1-4.5% (red, orange, dark blue and light blue lines) of the stained cell population as His-tag positive after infections with BacSfYB(His)₁₀. This was especially compared to the 40.3% His-tag staining of BacGFP(His)₁₀-infected cells. Single antibody stained cells infected with rBEVs functioned as negative controls for gating. Intracellular His-tag staining or infection efficiencies are presented as % of the stained cell population within scatter plots.

Therefore, the optimisation of rSfYB(His)₁₀ expression using rBEVs was restricted to Western blotting and BacSfYB(His)₁₀ vector clone 7. Monolayer Sf9 and Ao38 insect cell lines were infected with BacSfYB(His)₁₀, with Western blot analysis suggesting that the expression of

rSfYB(His)₁₀ was relatively equal between infected cell lines (Fig. 5.16A). This was especially the case given equal loading of lysates from infected insect cell lines was verified by Coomassie® Blue gel staining (Fig. 5.16B). Non-infected Sf9 cells (negative control for rSfYB(His)₁₀ expression) demonstrated no rSfYB(His)₁₀ expression when harvested in parallel (Fig. 5.16A), especially with relatively higher sample load as revealed by Coomassie® Blue gel staining (Fig. 5.16B). Furthermore, the expression dynamics of rSfYB(His)₁₀ protein was revealed to be optimally expressed at four or five days (96h or 120h) post-infection over a six day timecourse (Fig. 5.16C), given equal loading of samples by Coomassie® Blue gel staining (Fig. 5.16D). Additionally, non-infected Sf9 control cells were harvested in parallel at 1 and 6 days, and demonstrated no recombinant His-tag staining by Western blotting (Fig. 5.16C). Therefore, it was concluded that the 96h or 120h time points post-infection were the ideal harvest points for optimal rSfYB(His)₁₀ expression.

Despite optimising the point at which to harvest infected cells, the level of rSfYB(His)₁₀ expression in crude lysates was relatively less when compared to the more enhanced expression of rGFP(His)₁₀ (Fig. 5.16E). This was despite relatively equal loading between lysates as determined by Coomassie® Blue gel staining (Fig. 5.16F). It was also revealed that a significant Coomassie® Blue stained signal corresponding to rGFP(His)₁₀ protein was prominent compared to rSfYB(His)₁₀ in crude lysates (Fig. 5.16F, see red boxed, and compare with +BacSFYB(His)₁₀ lane). Instead, the expression of rSfYB(His)₁₀ in crude lysates derived from BacSFYB(His)₁₀-infected Sf9 resembled the protein profile of non-infected control (Fig. 5.16F, compare lanes NI and +BacSFYB(His)₁₀). This was further exemplified by Western blotting of elutes after His-tagged protein pulldown, and the relative difference in His-tagged recombinant protein pulldown was more pronounced (Fig. 5.16G, compare rGFP(His)₁₀ and rSfYB(His)₁₀ signals). Further to this, poor prep quality of pulldowns from either BacGFP(His)₁₀- or BacSFYB(His)₁₀- infected Sf9 lysates was concluded by Coomassie® Blue gel staining, with non-specific pulldown products observed (Fig. 5.16H, distinguish non-specific pulldown products from annotated His-tagged proteins).

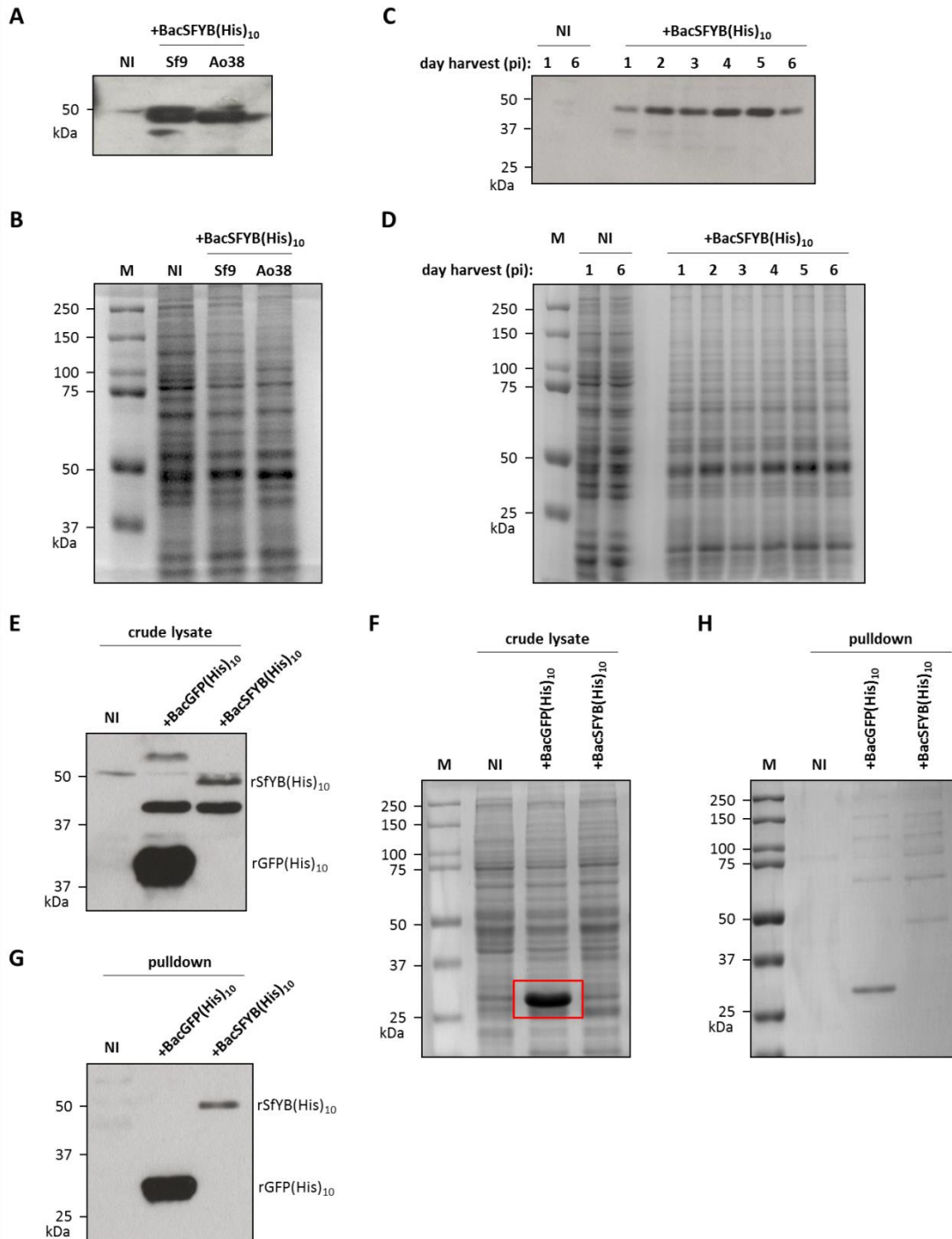


Figure 5.16 Optimisation of rSfYB(His)₁₀ expression for purification. Sf9 and Ao38 cells were infected with BacSFYB(His)₁₀, with non-infected Sf9 cells as a negative control, and **(A)** Western blot analysis demonstrated comparable expression of rSfYB(His)₁₀ between Sf9 and Ao38 cells using α 6His-HRP. Equal loading was demonstrated by **(B)** Coomassie[®] Blue gel staining, especially between infected Sf9 and Ao38 cells. **(C)** Western blotting revealed that the optimal harvest point for rSfYB(His)₁₀ was at day 4 or 5 post-infection with BacSFYB(His)₁₀ over a 6 day time course. Non-infected Sf9 cells were harvested at days 1 and 6 as negative controls. **(D)** Equal loading between negative control lanes or infected Sf9 lysates at each harvest interval was confirmed by Coomassie[®] Blue gel staining. **(E)** Western blot analysis showed His-tagged recombinant protein expression in crude lysates of Sf9 cells infected with BacGFP(His)₁₀ or BacSFYB(His)₁₀ 96h post-infection (see annotations for rSfYB(His)₁₀ and rGFP(His)₁₀). Non-infected Sf9 cells served as a negative control. Figure legend continues next page.

Figure 5.16 Optimisation of rSfYB(His)₁₀ expression for purification. F) Equal loading was demonstrated by Coomassie® Blue gel staining, with rGFP(His)₁₀ well distinguishable (see red boxed). **G)** Western blot of eluates of lysates from BacGFP(His)₁₀- or BacSfYB(His)₁₀- infected Sf9 using Dynabeads™ His-Tag Isolation and Pulldown, showed suboptimal pulldown of rSfYB(His)₁₀ was achieved relative to the pulldown of rGFP(His)₁₀. The prep qualities of pulldowns were considered poor by **(H)** Coomassie® Blue gel staining, with non-specific pulldown products evident. M, Precision Plus Protein™ Dual Colour Standards; NI, non-infected control; pi, post-infection.

The expression of rSfYB(His)₁₀ was initially found to be expressed at a more reasonable level relative to rGFP(His)₁₀ after infections of Sf9 cells by Western blot analysis (Fig. 5.17A). We then reasoned that the lysis protocols used to harvest rSfYB(His)₁₀ was not optimal. Therefore, rSfYB(His)₁₀ aa sequence was then assessed *in silico* and a hydrophobicity plot generated, which indicated that the rSfYB(His)₁₀ protein was largely hydrophilic with the given model – scorings for <0 was associated with hydrophilicity (Fig. 5.17A). Nonetheless, infected and non-infected Sf9 cells were fractionated for cytoplasmic and membrane fractions for Western blotting. This revealed that the target rSfYB(His)₁₀ was localised to the membrane fraction after α6His-HRP staining (Fig. 5.17B). Minimal levels of rSfYB(His)₁₀ was detected in the cytoplasmic fraction of infected Sf9 cells. This was in complete contrast to rGFP(His)₁₀ protein, which was instead localised in the cytoplasmic fraction and minimally detected in the membrane fraction of infected Sf9 cells (Fig. 5.17B). Sample loads were normalised by cell numbers, and as a result Coomassie® Blue gel staining revealed that, despite relatively equal loading between infected Sf9 fractions, rSfYB(His)₁₀ protein was easily distinguished with pronounced Coomassie® Blue staining (Fig. 5.17C, see red boxed). Similarly, rGFP(His)₁₀ was prominently stained in the cytoplasmic fraction by Coomassie® Blue staining (Fig. 5.17C, see blue boxed). The relevance of these findings was further substantiated by the lack of DNA contamination found in the cytoplasmic fractions of non-infected and infected Sf9 cells by agarose gel electrophoresis, with only ribosomal RNA species identified (Fig. 5.17D). Ultimately, the rSfYB(His)₁₀ protein was confirmed to be expressed in Sf9 cells using rBEVs, with optimal expression dynamics for the target His-tagged recombinant protein realised. Finally, the target protein was identified to localise with the membrane fraction of infected Sf9 cells, leading to an adjustment in harvest protocol.

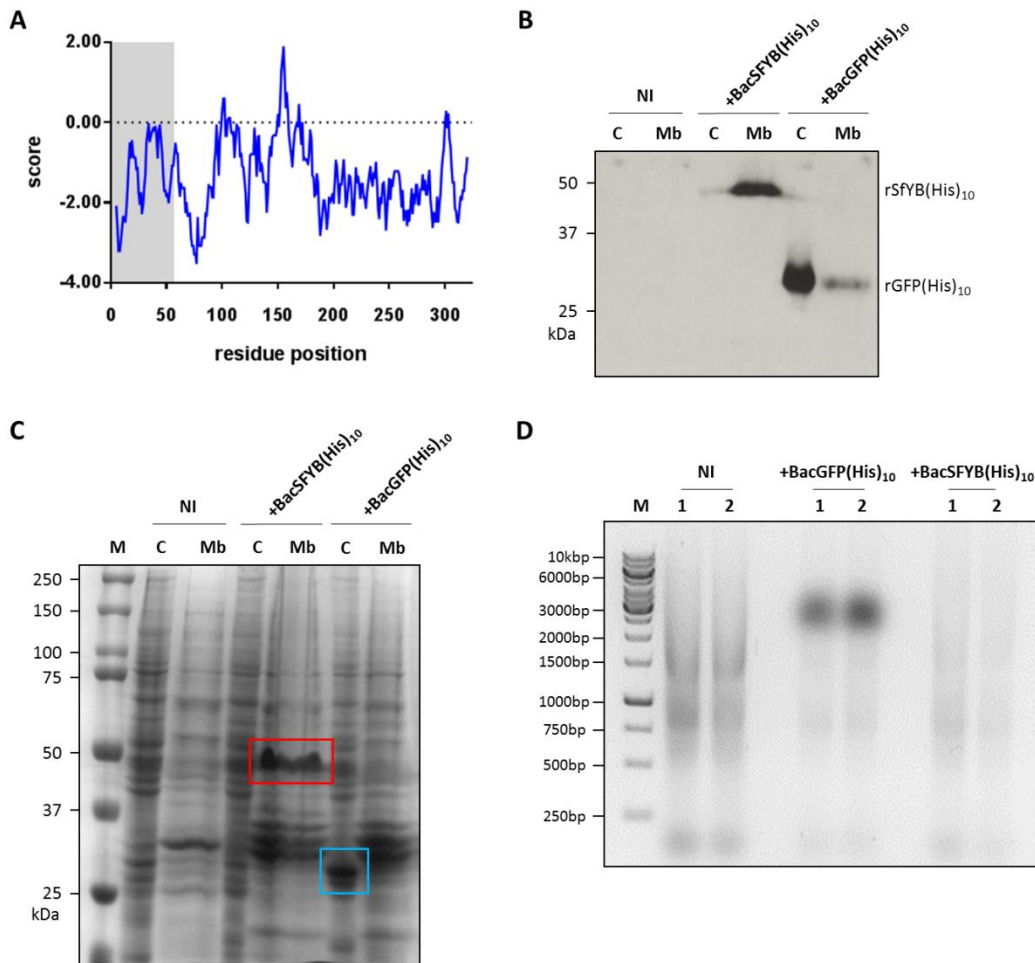


Figure 5.17 Recombinant SfYB(His)₁₀ expression localised in the membrane fraction of infected Sf9. A) Hydrophobicity plot of rSfYB(His)₁₀ protein was drawn by the ExpASy ProtScale programme using the Kyte and Doolittle scoring criteria (Kyte and Doolittle, 1982). This indicated that rSfYB(His)₁₀ is mostly hydrophilic with scores <0 along the length of the aa sequence. The y-axis indicates the hydrophobicity score for each residue position (x-axis; window size: 9), starting from the N-terminus. Grey shaded region corresponds to the His-tag sequence. Cytoplasmic and nuclear fractions were acquired of Sf9 cells infected with BacSFYB(His)₁₀ or BacGFP(His)₁₀, with non-infected Sf9 serving as negative control. Fractions were then analysed for expression and recombinant His-tagged protein localisation by **(B)** Western blotting using α6His-HRP, and revealed the target rSfYB(His)₁₀ was localised predominantly in the membrane fraction. Whereas rGFP(His)₁₀ was found in the cytoplasmic fraction (see annotations for both). **(C)** Coomassie® Blue gel staining demonstrated that loadings were well normalised by cell counting for cytoplasmic or nuclear fractions between infected and non-infected cells. The recombinant His-tagged protein was easily distinguished (see red and blue boxed annotations for rSfYB(His)₁₀ and rGFP(His)₁₀, respectively). Finally, cytoplasmic fractions were subjected to **(D)** 1% TAE agarose gel electrophoresis in duplicate for each infected and non-infected cells and shows that there was no genomic DNA contamination in harvested cytoplasmic fractions. C, cytoplasmic fraction; M, Precision Plus Protein™ Dual Colour Standards; Mb, membrane fraction; NI, non-infected control.

5.2.5 Purification of rSfYB(His)₁₀ for the development of antiserum against rSfYB(His)₁₀

An optimised harvest protocol for rSfYB(His)₁₀ involved sonication of infected Sf9 to disrupt DNA and membranes and include this fraction in whole lysates. Therefore, affinity

chromatography was implemented to immobilise rSfYB(His)₁₀ via its His-tag. The target recombinant protein was eluted in high Imidazole Elution Buffer, resulting in the identification of a peak absorbance reading by in-run spectrophotometry between elution fractions 3-11 (Fig. 5.18A). This likely corresponded to the elution of the target rSfYB(His)₁₀ protein, and was further confirmed by Coomassie® Blue gel staining of alternating fractions 5-21 (Fig. 5.18B). Initial Coomassie® Blue gel staining of the alternating fractions indicated that a predominant protein was eluted from infected Sf9 lysates, which resolved between 48-63kDa and was found mainly in early fractions 5-11 (Fig. 5.18A). The elution fraction number 5 showed additional non-specific elution products that resolved above the major 43-68kDa protein band, but relatively absent in remaining lanes. Therefore, sequential fractions between 6 and 15 were further assessed, wherein we identified the dominant eluted protein band resolving to approximately 50kDa (Fig. 5.18C). Furthermore, the tested elution fractions were considered relatively clean preps, given the absence of non-specific elution products (Fig. 5.18C). It was also noted that the rSfYB(His)₁₀ protein bound efficiently to the HisTrap Ni-Sepharose resin, as the flow-through after binding to the resin demonstrated no discernible signal that corresponded to the rSfYB(His)₁₀ protein (approximately 50kDa), whereas the corresponding protein band was present in lysates prior to loading the HisTrap column (Fig. 5.18C, compare lanes prior and FT). The 50kDa resolving protein likely corresponded to the target rSfYB(His)₁₀, but was nonetheless verified in the elution fractions 6-11 by Western blotting using α YB1 staining, with the dominant protein signal exhibited by fractions 6-8 (Fig. 5.18D).

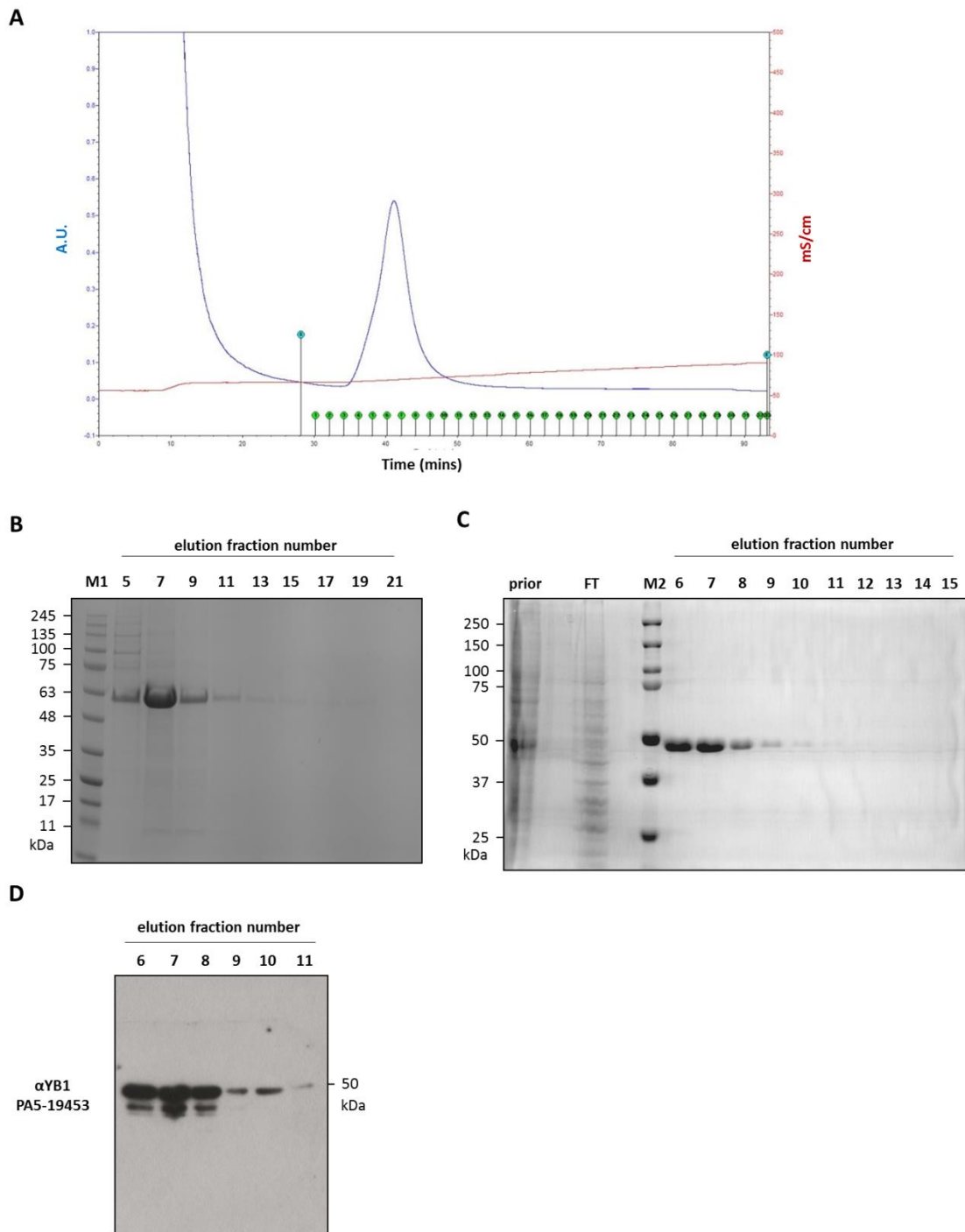


Figure 5.18 Purification of His-tagged rSFYB(His)₁₀ protein. Sf9 cells infected with BacSFYB(His)₁₀ were harvested, and lysates subjected to **(A)** affinity chromatography, which revealed that peak absorbance (A.U.) were recorded early for elution fractions 3-11 by in-run spectrophotometry. **(B)** Alternating fractions corresponding to 10 μ L elution fractions 5-21, and **(C)** 10 μ L all elution fractions 6-15, 50 μ L lysate before column loading (lane annotated prior) and 50 μ L flow-through (lane FT) after binding, were run on a 10% SDS-PAGE gels and Coomassie[®] Blue gel staining demonstrated that the rSFYB(His)₁₀ bound efficiently to the HisTrap column. This was mainly eluted in fractions 6 and 7 with good prep quality. **(D)** Purified rSFYB(His)₁₀ was also detected by Western blotting of elution fractions 6-11 using α YB1 (PA5-19453), which confirmed the specificity of the eluted product. FT, flow-through; M1, BLUEstain[™] Protein Ladder; M2, Precision Plus Protein[™] Dual Colour Standards; pre, lysate before affinity chromatography.

Purified rSfYB(His)₁₀ protein concentration was then calculated by densitometry (Fig. 5.19A). A sufficient quantity and concentration of rSfYB(His)₁₀ was thus purified – an average of 148ng/μL of rSfYB(His)₁₀ was calculated. Antiserum raised against the whole rSfYB(His)₁₀ was then produced (Covalabs), with antiserum from day 39 post-immunised rabbits (D39) tested by dot blot (Fig. 5.19B). Here, we show that the two-fold serial dilution of D39 antiserum from 1:4000-1:256,000 positively detected blotted rSfYB(His)₁₀ protein in a dose-dependent manner (Fig. 5.19B). Additionally, the specificity of the D39 antiserum towards endogenous SfYB protein was confirmed by staining 10μg of Sf9 (non-infected with BacSfYB(His)₁₀) and even 293T and YB1 knockout lysates with D39 antiserum. Antiserum reactivity was observed at 1:4000 and 1:32,000 dilutions for Sf9 and 293T lysates. The detection of endogenous SfYB protein using D39 antiserum was more pronounced using Sf9 lysate compared to 293T lysate at 1:32,000 dilution, which was perhaps due to the reduced homology observed between SfYB and human YB1. Signal from YB1 knockout lysate was completely absent at the given dilution factor also, but staining of YB1 knockout lysate using 1:4000 dilution was capable of producing a detectable signal. Given that YB1 knockout has been previously confirmed using αYB1 (see Chapter 3), the detected signal at 1:4000 dilution factor was possibly a product of non-specific binding given perhaps too low a D39 antiserum dilution was used. The above was further substantiated with control staining using only D39 antiserum or HRP-conjugated secondary antibodies showing no background signal. Especially, the complete staining of blotted rSfYB(His)₁₀ using antiserum derived from pre-immunisation (day 0; D0) at 1:4000 dilution and HRP-conjugated secondary antibodies did not yield detectable signal either, implying that the pre-immunised rabbits did not harbour a pre-existing pool of antibodies targeting rSfYB(His)₁₀. Positive control staining using α6His-HRP primary antibody was capable of detecting blotted rSfYB(His)₁₀, but only at 1:4000 and 1:32,000 dilutions.

The specificity of D39 antiserum to endogenous SfYB and rSfYB(His)₁₀ was also demonstrated by Western blotting (Fig. 5.19C). Here the approximately 50kDa protein corresponding to rSfYB(His)₁₀ or endogenous SfYB could be stained using D39 antiserum at 1:32,000 dilution (Fig. 5.19C, top panel, see rSfYB(His)₁₀ 200ng and Sf9 lanes, respectively). Additionally, staining using 6His-HRP only picked up the His-tagged recombinant protein and not endogenous SfYB (Fig. 5.19C, bottom panel, compare rSfYB(His)₁₀ 200ng lane to Sf9 lane). Protein loads were also verified by Coomassie® Blue gel staining, where rSfYB(His)₁₀ and Sf9 lysate was only loaded in their corresponding lanes (Fig. 5.19D). Altogether, Western blotting using D39 antiserum and α6His-HRP differentiated that the D39 antiserum was specific for SfYB.

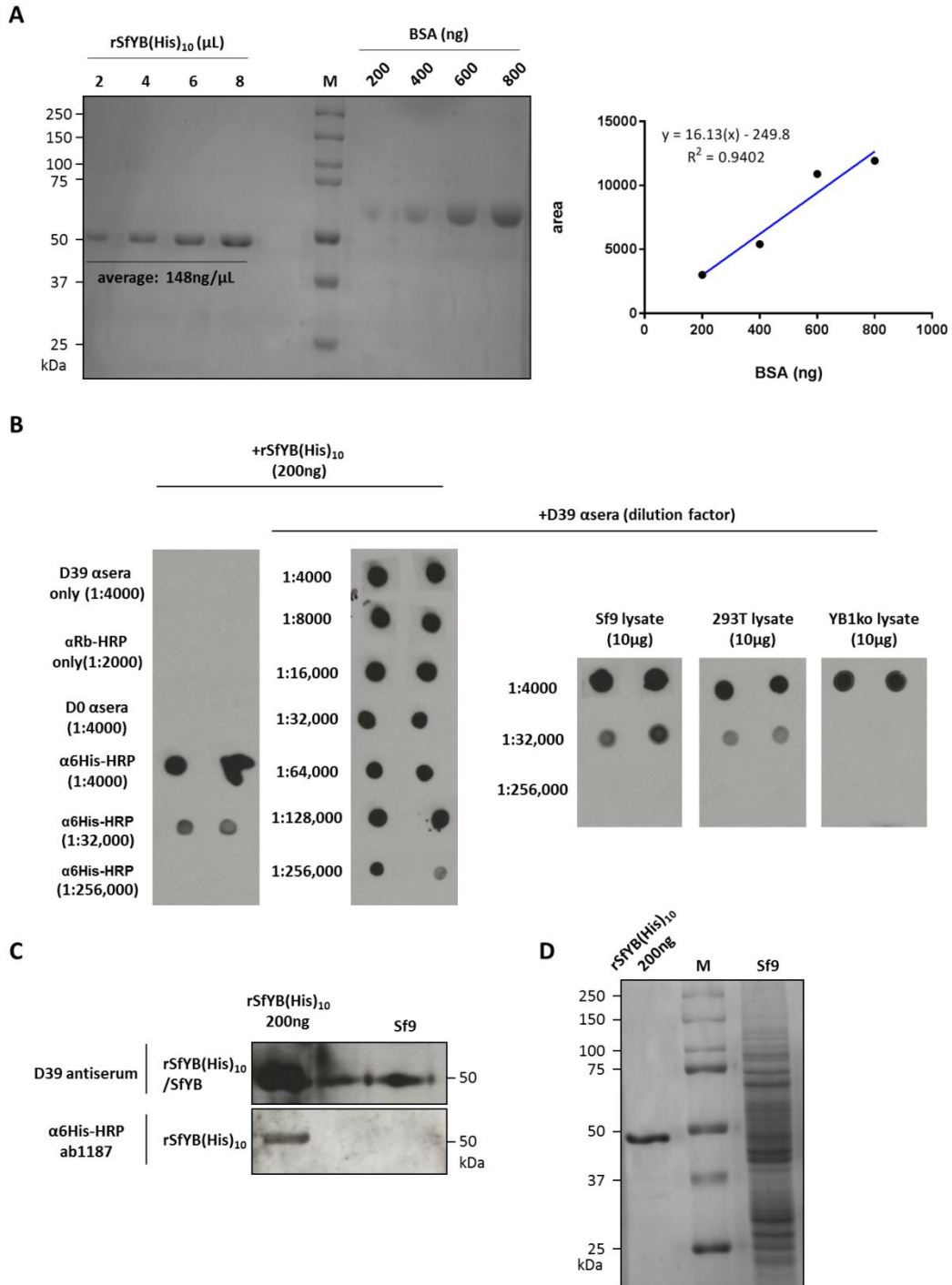


Figure 5.19 Antiserum raised against rSfYB(His)₁₀ shows specificity for SfYB. **A)** Coomassie[®] Blue gel staining of known quantities of BSA and known volumes of rSfYB(His)₁₀ was used generate of a standard curve. An average 148ng/µL for rSfYB(His)₁₀ was calculated by densitometry. **B)** Dot blot assessment of D0 (pre-immunisation) and D39 (post-immunisation) antiserum raised against rSfYB(His)₁₀. Dot blot revealed reactivity of D39 antiserum to rSfYB(His)₁₀ and endogenous SfYB protein, but to a lesser extent human YB1. Staining of rSfYB(His)₁₀ with only D39 antiserum or secondary, and complete staining with D0 antiserum served as negative controls. Complete staining of rSfYB(His)₁₀ protein using α6His-HRP served as positive controls. **C)** D39 antiserum was able to specifically detect rSfYB(His)₁₀ and endogenous SfYB in Sf9 lysate by Western blotting (top panel), with α6His-HRP distinguishing rSfYB(His)₁₀ from endogenous (bottom panel). **D)** Protein loads of rSfYB(His)₁₀ and Sf9 lysates were

confirmed by Coomassie® Blue gel staining. BSA, bovine serum albumin; D0, day 0 antiserum; D39, day 39 antiserum; M, Precision Plus Protein™ Dual Colour Standards; YB1ko, YB1 knockout.

5.3.0 Chapter summary

With the use of a multitude of *in silico* analyses, we deduced the aa sequences and predicted 3D-structures (particularly of the CSD) of previously uncharacterised Y-Box proteins of *Spodoptera* spp. (*S. frugiperda*, *S. exigua*, *S. litura*, and *S. littoralis*) and *T. ni*. High conservation of the whole molecules was identified, especially with respect to the defining CSD regions (almost 100%). This was even the case between Y-Box proteins and human YB1. Molecular cloning and expression of recombinant His-tagged SfYB in its native host cell line (Sf9) was conducive, and perhaps even advantageous compared to expression in *T. ni* based cell lines. And perhaps represents the first attempt to express and purify a recombinant protein derived from *S. frugiperda* or Sf9 using baculovirus technology. This was followed by the successful generation of antiserum that was specific for the target endogenous protein from Sf9 lysates.

The Sf9 cell line is robustly used for the propagation of rAAV vectors, but the interactions or impact of host cell proteins on rAAV vector processing has not been characterised. In having identified a Y-Box protein for *S. frugiperda*, and raised a specific antiserum for endogenous SfYB, we can examine the impact of SfYB protein on rAAV vector processing. Therefore, in the next chapter, we aim to target SfYB protein expression for gene knockout and explore the effects on rAAV vector production.

Chapter 6: Generation of SfYB Knockout Cell Lines for Recombinant AAV Vector Production Using CRISPR/Cas9 Genome Editing

6.1.0 Introduction

Sf9 cells represents perhaps one of the most clinically relevant models for rAAV(1) vector production (Carpentier *et al.*, 2012). In the present study, we utilised the knowledge that an existing Y-Box protein homologue (SfYB) is expressed by Sf9 (Chapter 5), and that the knockdown of human YB1 in 293T cells correlated with an enhanced rAAV vector producer cell line (Satkunanathan *et al.*, 2014). We therefore aimed to modify SfYB protein expression in Sf9 cells using CRISPR/Cas9 genome editing, and characterised these cells to show that targeted disruption of SfYB protein expression was successfully achieved. This was supported by analyses of both phenotype and genotype. We also show the significant advantages of establishing single cell clones after CRISPR/Cas9 genome editing, and screening for protein knockout using specific antibodies that was raised against the target protein of interest. However, ultimately, we also show that the disruption of SfYB expression in Sf9 cells did not confer an enhanced rAAV vector producer cell line.

6.2.0 Results

6.2.1 Designing and cloning SFYB-specific gRNAs into CRISPR plasmids

We previously identified Y-Box protein homologues for *Spodoptera spp.* including *S. frugiperda* (SfYB; see Chapter 5); from which ovarian tissue established the Sf9 cell line (Vaughn *et al.*, 1977). In order to design gRNAs to target the gene encoding SfYB (*SFYB*) we first identified the gene from which the BYB Y-Box protein orthologue was expressed from in *B. mori*, *in silico*. The *BYB* gene was annotated for exon and intron sequences based on the genomic scaffold (NW_004582015.1) and available transcriptomics data as per GCF_000151625.1 (Fig. 6.1A). Initially, the availability of assembled genomic scaffolds for *Spodoptera spp.* was limited to *S. litura*; for which, an evolutionary conserved Y-Box protein homologue was identified *in silico* (NC_036217.1, LOC111360813) (Cheng *et al.*, 2017b). Therefore, the *SLYB* sequence was mapped for exon and intron sequences based on available

transcriptomics data as per LOC11360813 (Fig. 6.1B). Given the significant homology and evolutionary conservation between Y-Box protein homologues (section 5.3.2) we reasoned that the gene structures would be conserved, also. Using both Y-Box protein homologue gene annotations as genomic scaffolds *SFYB*'s exons (1-5) were mapped for the assembled contig accession OE0A01010394.1 (Fig. 6.1C). Following which, gRNAs (gRNAsf1-3) specific for exonic sequences were designed manually, which encompassed a 20nt sequence that preceded the 5'-NGG-3' PAM sequence (Fig. 6.1C, annotated gRNAsf1-3 designs are depicted relative to their target location in exon 1-3, respectively) for cloning into the GeneArt™ linearised, all-in-one CRISPR nuclease vector.

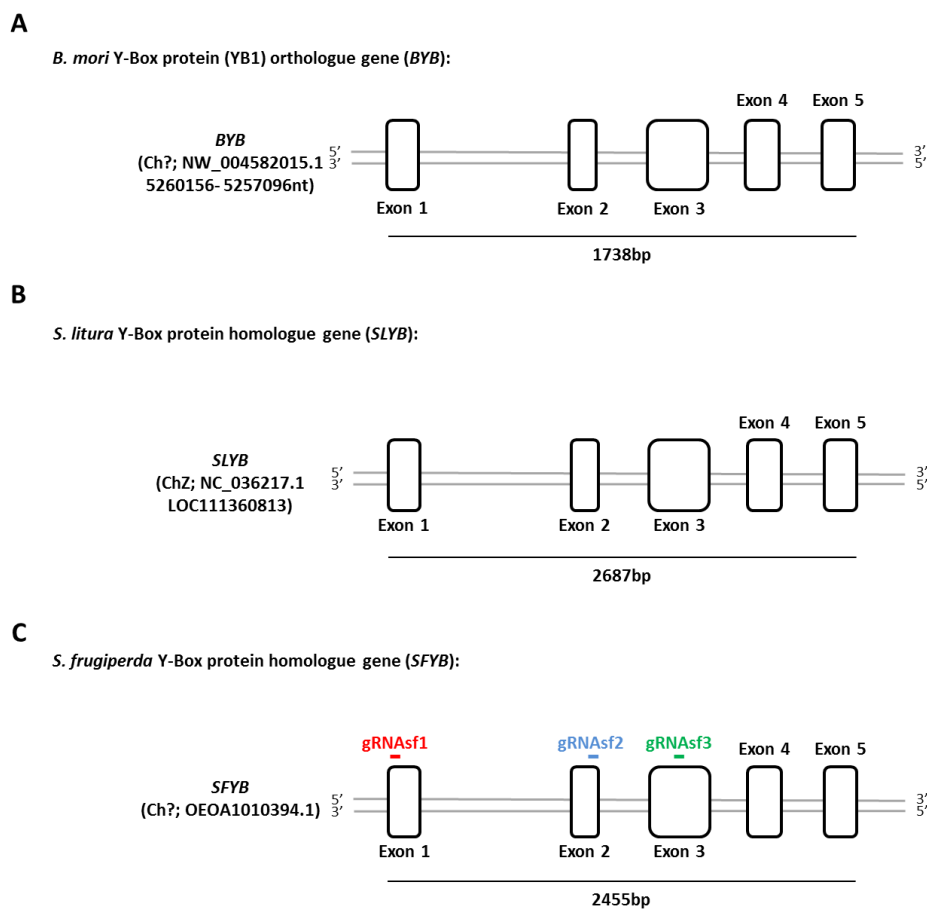


Figure 6.1 Inferring *SFYB* gene exon and intron sequences using *BYB* and *SLYB* genes as genomic scaffolds. The Y-Box protein gene homologues endogenous **(A)** to *B. mori* (*BYB*) and **(B)** *S. litura* (*SLYB*) were mapped for exon and intron sequences based on the assembled sequences and available transcriptomics data. These were mapped to show 5 exons each, and were used as genomic scaffolds to map **(C)** the exon sequences of *S. frugiperda*'s Y-Box protein gene homologue (*SFYB*) based on DNA sequence homology using the OE0A01010394.1 assembled contig. Designed gRNAs (gRNAsf1-3) are annotated to depict their relative target sequences in exon 1-3, respectively.

The strategy employed to clone gRNAsf1-3 DNA oligos into the GeneArt™ linearised, all-in-one CRISPR nuclease vector is depicted in Fig. 6.2A. Cloning of each pCRISPR plasmids (pCRISPR-

SFYBsgRNAsf1-3) was confirmed by Sanger sequencing (Figs. 6.2B-C). Here the DNA sequences corresponding to gRNAsf1-3 were correctly identified (Figs. 6.2B-C, see underlined red, blue, and green sequences for gRNAsf1-3, respectively), for which flanking sequences corresponding to the GeneArt™ all-in-one CRISPR nuclease vector backbone. Therefore, it was deduced that all three pCRISPR plasmids encoded the desired gRNA DNA sequences specific for the *SFYB* gene under U6 promoter control.

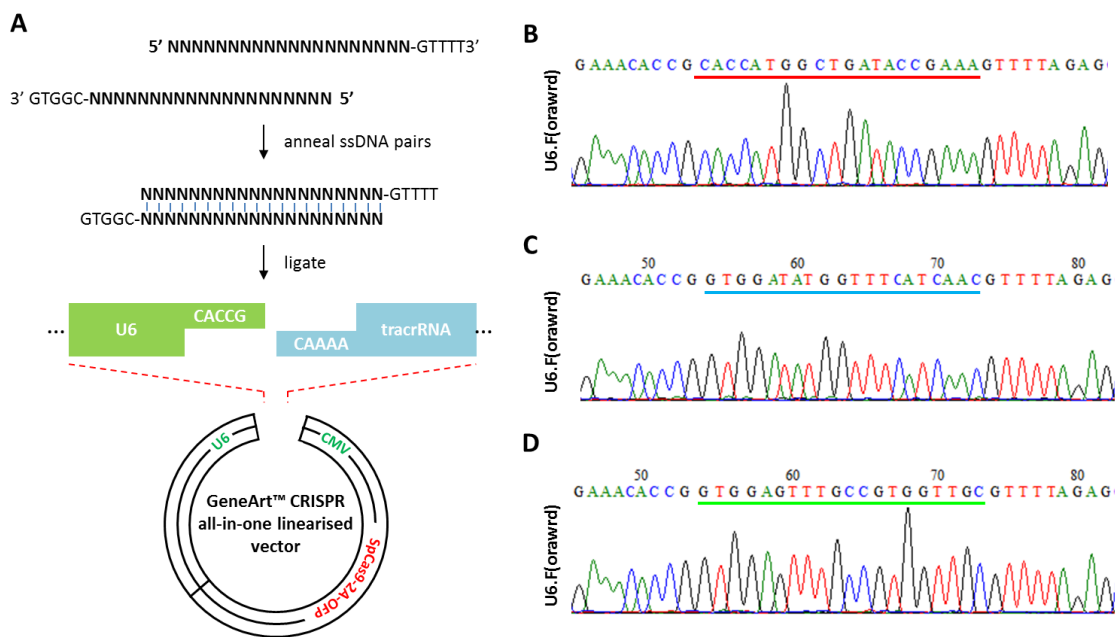


Figure 6.2 Cloning of gRNAsf1-3 oligos into GeneArt™ linear, all-in-one nuclease vector. **A)** Schematic of strategy employed to clone gRNA DNA (gRNAsf1-3) designs into GeneArt™ linear, all-in-one nuclease vector for construction of pCRISPR-*SFYBsgRNAsf1-3* plasmids, respectively. Cloning was confirmed by Sanger sequencing using the U6.F(oward) sequencing oligo for **(B)** pCRISPR-*SFYBsgRNAsf1*, **(C)** pCRISPR-*SFYBsgRNAsf2*, and **(D)** pCRISPR-*SFYBsgRNAsf3* plasmids, with gRNA DNA sequences correctly verified (see red, blue, and green underlined sequences, respectively for gRNAsf1-3).

6.2.2 Optimising transfections for targeted knockout of SfyB protein expression

The pCRISPR plasmids encode the SpCas9-2A-OFP ORF under CMV promoter control. Therefore, transfection of Sf9 cells with pCRISPR plasmids and the expression of OFP was examined 72h post-transfection by fluorescence microscopy and flow cytometry. Transfection of Sf9 cells with pCRISPR-*SFYBsgRNAsf1-3* resulted in only a few cells per field of view expressing OFP (Fig. 6.3A, panels a and b, c and d, and e and f, respectively). This was similar to mock transfected Sf9 cells, that exhibited no fluorescent signal indicative of GFP or OFP expression (Fig. 6.3A, panels g-h). However, positive transfection control with pIEx-1.eGFP was deemed sufficient with notable GFP expression driven by IE1 promoter (Fig. 6.3A, panels j and k). The above was further recapitulated by flow cytometry analysis, where transfection of

8.5x10⁵ Sf9 cells with 0.5µg or 3µg of pCRISPR plasmids demonstrated no detectable GFP-positive cell populations 72h post-transfection, and were considered comparable to 0.25% exhibited by mock transfected control (Fig. 6.3B). Sf9 cells transfected with pIEx-1.eGFP were detectable even with the PE channel (Fig. 6.3B, 48.6% for 0.5µg input). This was due to the broad wavelength spectrum emitted by GFP excited from the 488nm laser and is consequently picked up in the PE channel, and also because GFP was expressed under the insect-specific IE1 promoter control. This was in spite of the fact that approximately 59.3% Sf9 cells were GFP-positive, 72h post-transfection with 0.5µg pIEx-1.eGFP of 8.5x10⁵ cells (Fig. 6.3C, see green line). Expectantly, the pCRISPR-transfected cells exhibited negligible GFP fluorescence comparable to mock transfection control (FFig. 6.3C).

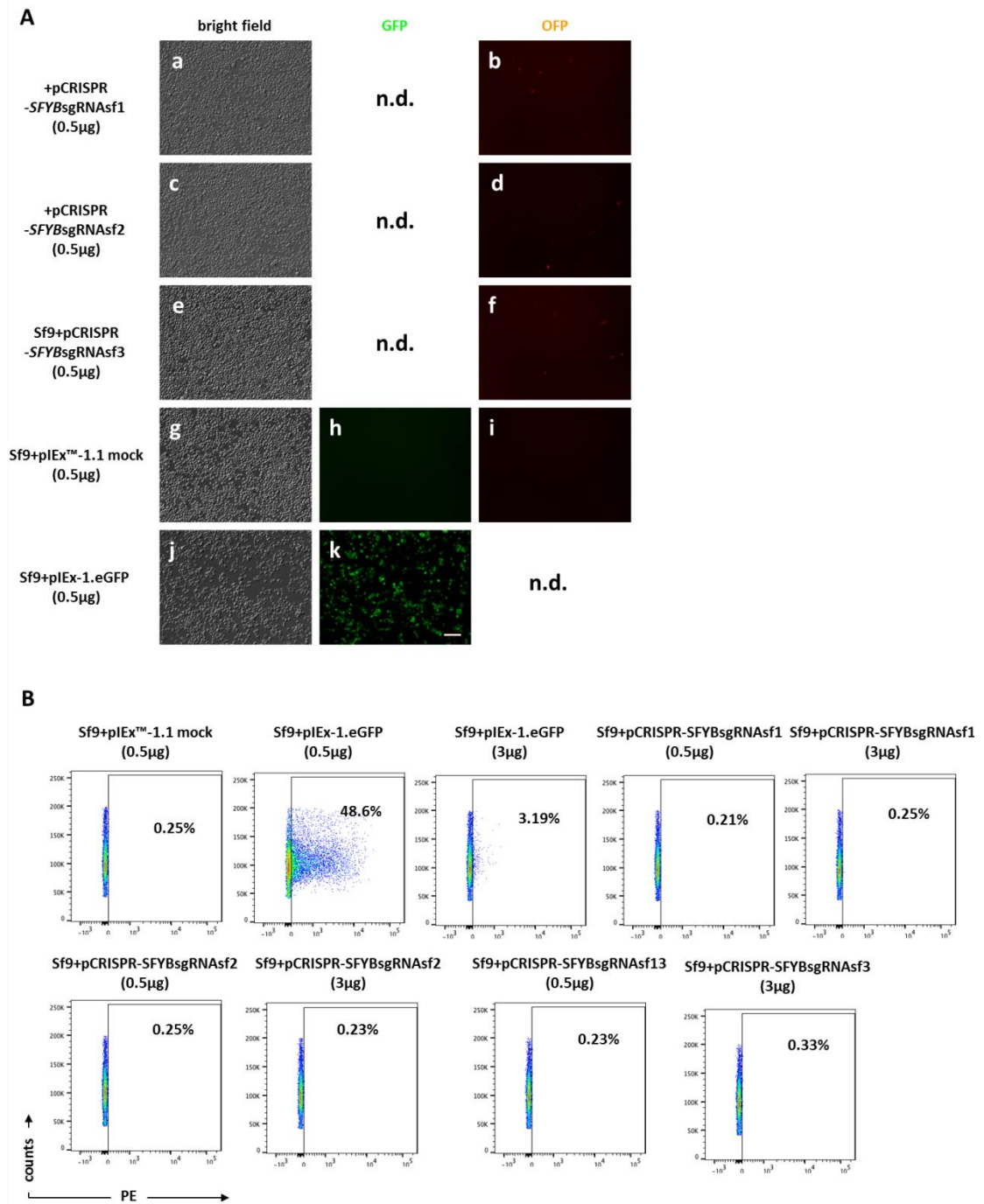


Figure 6.3 Transfection of Sf9 cells with pCRISPR-SFYBsgRNAsf1-3. Figure and Figure legend continues next page.

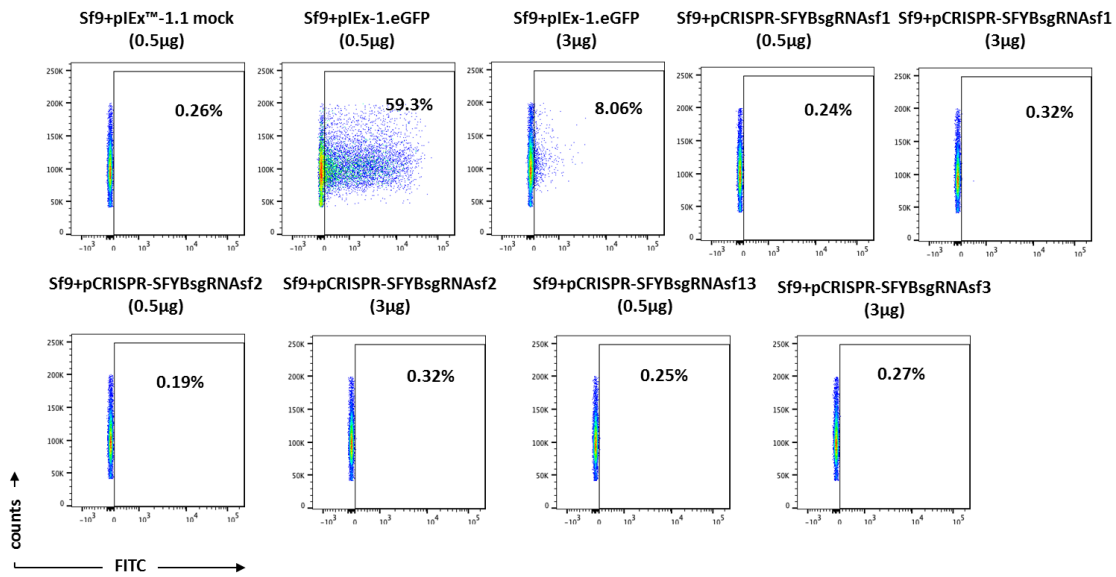
C

Figure 6.3 Transfection of Sf9 cells with pCRISPR-SFYBsgRNAsf1-3. Transfected Sf9 cells were analysed by bright field and fluorescence microscopy 72h post-transfection using CellfectinTM II. **A)** Low to negligible OFP expression was exhibited by Sf9 cells transfected with pCRISPR-SFYBsgRNAsf1-3 (panels a and b, c and d, and e and f, respectively). Transfection with pIEx-1.1 served as mock transfection control, and demonstrated no fluorescence signal for GFP or OFP (panels g-i), or pIEx-1.eGFP (positive transfection control) demonstrated positive GFP expressing populations (panels j and k). Representative scale bar = 50μm for all panels. Plasmid pCRISPR transfection efficiency was also examined by flow cytometry **(B)** for OFP expression, of which fluorescence was comparable to mock transfection control (black line), despite **(C)** adequate transfection efficiency demonstrated by 0.5μg pIEx-1.eGFP reporter plasmid (59.3%). The % of fluorescent cells is displayed within scatter plots. GFP, green fluorescent protein; n.d., not determined; OFP, orange fluorescent protein.

Due to the inefficiency of pCRISPR-SFYBsgRNAsf1-3's CMV promoter to express SpCas9-2A-OFP in Sf9 cells, instead, we opted to transfect Sf9 cells with sgRNA:recombinant Cas9 ribonucleoprotein (RNP) complexes. However, transfection with RNPs does not include a reporter system to track transfection efficiency. Additionally, CRISPRmaxTM-mediated transfection of Sf9 or insect cells was not previously described by the manufacturer. To ensure Sf9 cells could be efficiently transfected with nucleic acid using CRISPRmaxTM Reagent, pIEx-1.eGFP reporter plasmid was used to examine transfection capacity of the transfection reagent. Additionally, we cross-examined the effect of different serum-free media as diluents on transfection efficiency of nucleic acids (Fig. 6.4). We observed a superior transfection capacity of Sf9 cells for pIEx-1.eGFP plasmid when transfection mixes were prepared with opti-MEMTM as diluent (Fig. 6.4A). This was in contrast to mixes prepared with Sf-900TM II SFM or Grace's Insect Medium, across the range of pIEx-1.eGFP input quantities tested (Fig. 6.4B-C). The transfection of 4×10^5 Sf9 cells with 20μg of pIEx-1.eGFP corresponded to the approximate amount of input quantity (pmol) of 125ng of sgRNA and 500ng of recombinant Cas9 protein

recommended by the GeneArt™ Platinum™ Cas9 Nuclease Kit (Invitrogen). Transfection efficiencies of approximately 98-99% was achieved using opti-MEM™ compared to the 29.3% or 46.3% transfection efficiencies exhibited by using Sf-900™ II SFM and Grace's Insect Medium for 20µg plasmid quantity, respectively (Fig. 6.4A-C). Mock transfected Sf9 cells across all transfection mix preparation methods also demonstrated negligible fluorescence of 0.27-0.36% (Fig. 6.4A-C). Therefore, it was reasonably deduced that opti-MEM™ functioned as the optimum diluent for the transfection of sgRNA:recombinant Cas9 RNP complexes using CRISPRmax™ Reagent.

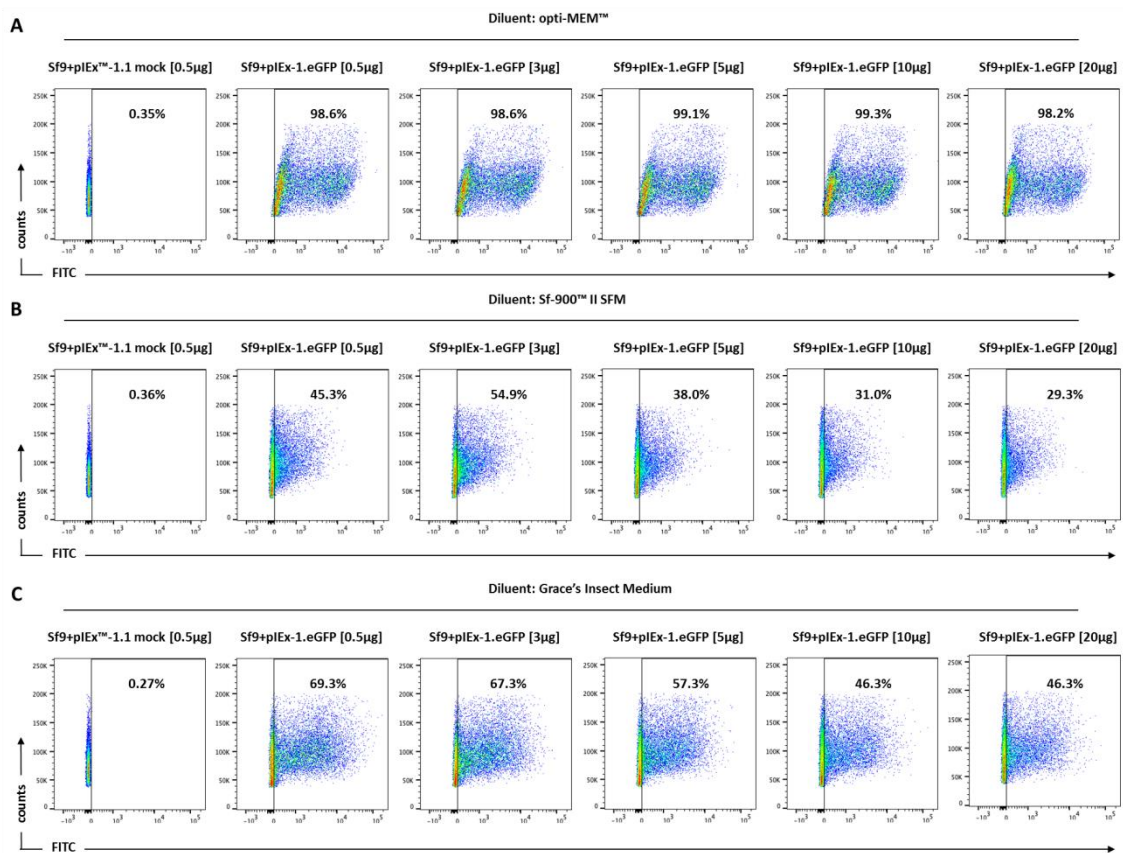


Figure 6.4 Optimisation of Sf9 transfections using CRISPRmax™ Reagent and different diluents by flow cytometry analysis. Sf9 cells were mock transfected with pEx-1.1 (negative control) and pEx-1.eGFP plasmid prepared in either Sf-900™ II SFM, Grace's Insect Medium, or opti-MEM™, and analysed by flow cytometry. Flow cytometry revealed that **(A)** opti-MEM™ functioned as the optimum diluent to prepare transfection mixes, and showed almost 100% transfection efficiency for all input plasmid quantities (0.5, 3, 5, 10, and 20µg) when compared to mock. This was compared to modest transfection efficiencies ranging between 29.3-69.3%, which was achieved using **(B)** Sf-900™ II SFM or **(C)** Grace's Insect Medium as diluent for input plasmid quantities (0.5, 3, 5, 10, and 20µg). Transfection efficiencies are presented as % within scatterplots.

The sgRNAs targeting *SFYB* gene at the putative *SFYB* exons 1, 2 and 3, were synthesised by *in vitro* transcription (IVT) (Fig. 6.5). Assembly PCR of each sgRNAsf1-3 DNA template were PCR amplified and verified by agarose gel electrophoresis as approximately 125bp PCR amplicons

(Fig. 6.5A). Subsequent IVT reactions resulted in the synthesis of the sgRNAsf1-3 recognised as approximately 100nt in length, and distinct from input sgRNAsf1-3 DNA templates (approximately 125bp), by agarose gel electrophoresis (Fig. 6.5B, compare annotated sgRNA bands and DNA template bands). Finally, DNase I treated and purified sgRNAsf1-3 was further confirmed by agarose gel electrophoresis as isolated 100nt sgRNA species, with no detectable DNA template as contaminants (Fig. 6.5C).

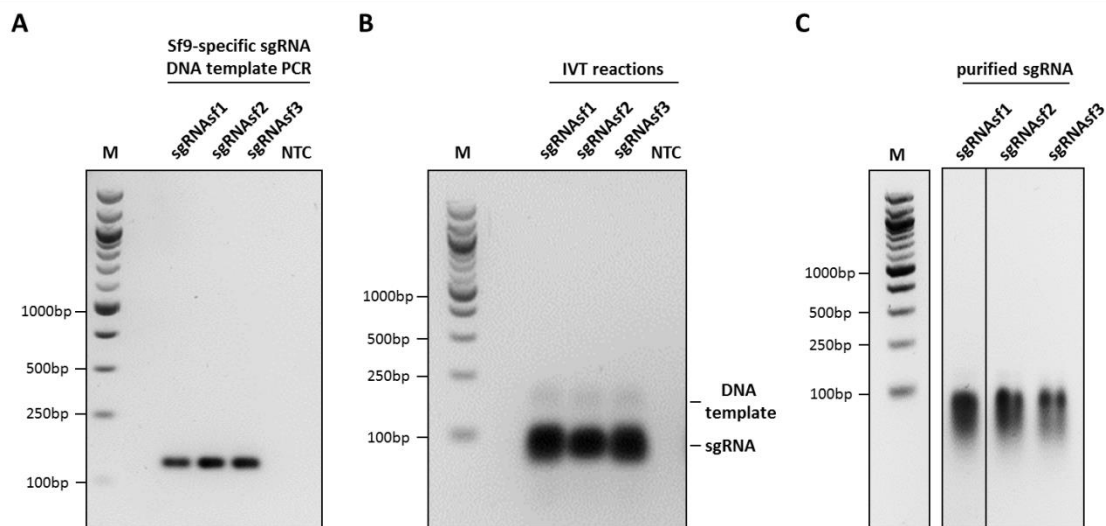


Figure 6.5 IVT reactions and purification of sgRNAs for *SFYB* targeting. **A)** 2% TBE agarose gel electrophoresis of sgRNAsf1-3 DNA template PCR by assembly PCR using the GeneArt™ Precision gRNA Synthesis Kit, which demonstrated that approximately 125bp DNA templates for each sgRNA design was successfully amplified, especially when compared to assembly PCR of non-template control reaction (Tracr fragment plus universal T7 primer mix was omitted). **B)** 2% TBE agarose gel electrophoresis of IVT reactions for successful synthesis of sgRNAsf1-3 as approximately 100nt ssRNA (bands annotated sgRNA), relative to the input 125bp DNA template (annotated) and non-template control reaction (DNA template from non-template control was added for IVT reaction). **C)** 2% TBE agarose gel electrophoresis of DNase I-treated and purified sgRNA as isolated 100nt (approximately) ssRNA species corresponding to the target sgRNAsf1-3. IVT, *in vitro* transcription M, 100bp DNA Ladder; NTC, non-template control.

For *SFYB* gene knockout, Sf9 cells were transfected with sgRNA (sgRNAsf1-3):recombinant Cas9 RNP complexes. Single-cell cloning of transfected Sf9 cells was attempted 72h post-transfection; however, Sf9 cells did not survive at low densities as listed in Tables 6.1 and 6.2. Parental Sf9 and transfected Sf9 cells showed no cell growth after one month in culture at seeding densities of 0.3, 10, 200, and 500 cells/well in 96-well plates using complete Sf900™ II SFM (Table 6.1, red highlighted cells). Variation in media composition, i.e. Sf-900™ II SFM with/without 5% FCS, complete Sf-900™ II SFM with/without 5% FCS, or conditioned media, did not improve cell growth at very low cell densities (Table 6.2, red highlighted cells). Higher cell densities of 1×10^4 and 5×10^3 cells/well of a 96-well plate were required for cell growth (Tables 6.1 and 6.2, blue highlighted cells), or confluency by one week in culture at the seeding

density of 1×10^3 cells/well (Tables 6.1 and 6.2, orange highlighted cells). Therefore, establishing a cell line from a single cell was not practical.

Table 6.1 Observations from single-cell cloning of Sf9 cells 72h post-transfection with sgRNAs and recombinant Cas9

	Sf9 non-transfected	Sf9+sgRNAsf1:rCas9 RNP	Sf9+sgRNAsf2:rCas9 RNP	Sf9+sgRNAsf2:rCas9 RNP
	growth media variant:			
cell seeding density/well	complete Sf-900™ II SFM	complete Sf-900™ II SFM	complete Sf-900™ II SFM	complete Sf-900™ II SFM
1.00E+04	confluent	confluent	confluent	confluent
5.00E+03	confluent	confluent	confluent	confluent
1.00E+03	confluent	confluent	confluent	confluent
5.00E+02	ng	ng	ng	ng
2.00E+02	ng	ng	ng	ng
1.00E+02	ng	ng	ng	ng
1.00E+01	ng	ng	ng	ng
3.00E-01	ng	ng	ng	ng

ng, no growth was detected by 1 month (red highlighted cells) post-seed; blue highlighted cells indicate confluency by 72h or by one week (orange highlighted cells); rCas9, recombinant Cas9.

Table 6.2 Observations from single-cell cloning of Sf9 cells with various Sf900™ II SFM variations

cell seeding density/well	Sf9			
	growth media variant			
	Sf900™ II SFM	Sf900™ II SFM + 5% FCS	complete Sf900™ II SFM + 5% FCS	conditioned media
1.00E+04	confluent	confluent	confluent	confluent
5.00E+03	confluent	confluent	confluent	confluent
1.00E+03	confluent	confluent	confluent	confluent
5.00E+02	ng	ng	ng	ng
2.00E+02	ng	ng	ng	ng
1.00E+02	ng	ng	ng	ng
1.00E+01	ng	ng	ng	ng
3.00E-01	ng	ng	ng	ng

ng, no growth (red highlighted cells) was observed by 1 month post-seed; blue highlighted cells indicate confluency by 72h or by one week (orange highlighted cells); rCas9, recombinant Cas9.

6.2.3 Characterising Sf9 cells for SfYB knockout

The strategy to establish SfYB knockout cells is illustrated (Fig. 6.6). This involved multi-runs of CRISPR RNP transfection to off-set the potential that only a sub-population of cells were transfected with sgRNA(sf1-3):recombinant Cas9 RNP complex at a ratio of 1X and 10X of the sgRNA and recombinant Cas9 to CRISPRmax™ Reagent.

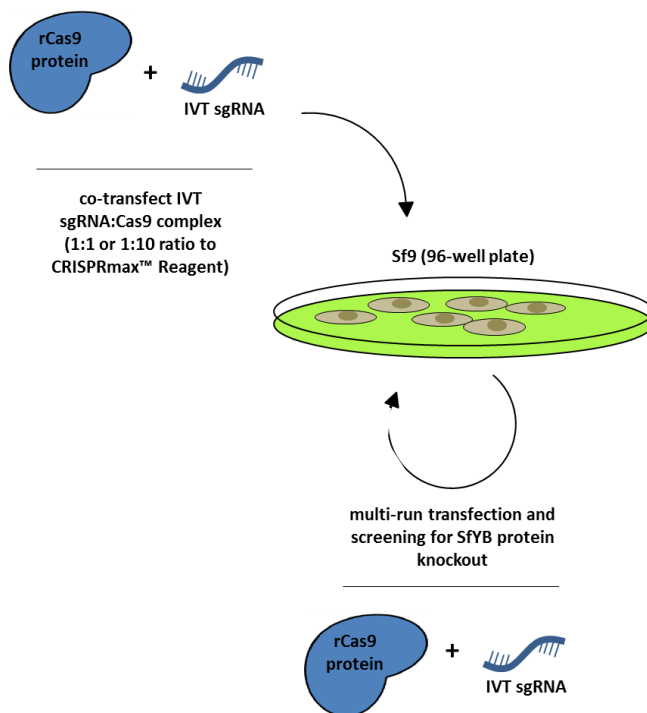


Figure 6.6 Strategy employed to transfect (multi-run transfection) Sf9 cells with sgRNA and recombinant Cas9 complex for targeted disruption of *SFYB* gene. Sf9 cells were transfected with 1X or 10X more sgRNA(sf1-3) and recombinant Cas9 relative to CRISPRmax™ Reagent, to off-set transfection efficiency limitation per transfection cycle. IVT, *in vitro* transcription; rCas9, recombinant Cas9.

Initial Western blotting for SfYB protein expression in bulk transfected cells demonstrated notable disruption of SfYB expression by sgRNAsf1, which targeted *SFYB*'s putative exon 1 (Fig. 6.7). 293T cell lysate also functioned as a positive control for YB1/Y-Box protein homologue staining with α YB1, given the unavailability of the antiserum raised against rSfYB(His)₁₀ at the time. Western blot analysis of Sf9 cells transfected with sgRNAsf -2 and -3 (targeting putative exon 2 and 3, respectively) depicted a minimal or lack of genome editing of the *SFYB* gene (Figs. 6.7B and C, respectively). Coomassie® Blue gel staining was used to demonstrate equal loading (Figs. 6.7D-F for *SFYB* exon 1-3 targeting, respectively).

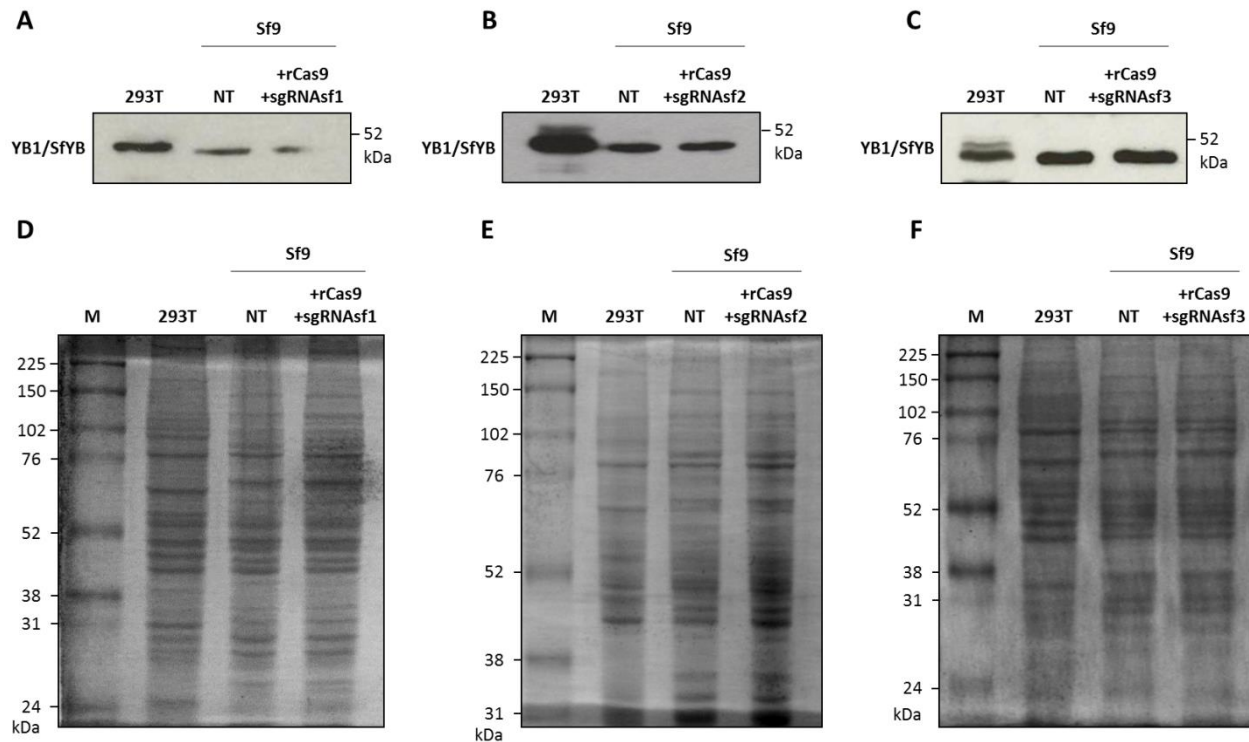


Figure 6.7 Establishing SfYB knockout cell lines. Western blot analysis using α YB1 (PA5-19453) of bulk transfected Sf9 cells with sgRNA and recombinant Cas9 (1:1 ratio with CRISPRmax™) was compared to non-transfected Sf9 and 293T controls for CRISPR/Cas9-mediated disruption of the *SFYB* gene by targeting (A) exon 1 with sgRNAsf1, with SfYB expression seemingly disrupted. Whereas, targeting (B) exon 2 using sgRNAsf2, and (C) exon 3 using sgRNAsf3 both exhibited no disruption in the expression levels of SfYB when compared to non-transfected Sf9 controls. Coomassie® gel staining demonstrated relatively equal loads between Sf9 non-transfected control and transfected Sf9 cell lysates, for CRISPR/Cas9 targeting of (D) exon 1, (E) exon 2, and (F) exon 3. M, ECL™ Rainbow™ Marker – Full Range; NT, non-transfected control; rCas9, recombinant Cas9.

Over seven multi-rounds of transfection were performed for sgRNAsf1, with Western blotting analysis indicating a progressively increased disruption of the SfYB protein expression since the first round (Figs. 6.8A and B). This was especially the case of Sf9 cells consecutively transfected with 10:1 ratio of sgRNAsf1 and recombinant Cas9 RNP, where a diminished SfYB protein signal was evident compared to non-transfected Sf9 control (Figs. 6.8A and B). Further transfections at the 10:1 ratio eventuated in the complete disruption of SfYB expression at the sixth and seventh round of transfection, especially when compared to non-transfected Sf9 control (Fig. 6.8B). Equal protein loads was determined by Coomassie® Blue gel staining (Figs. 6.8C and D).

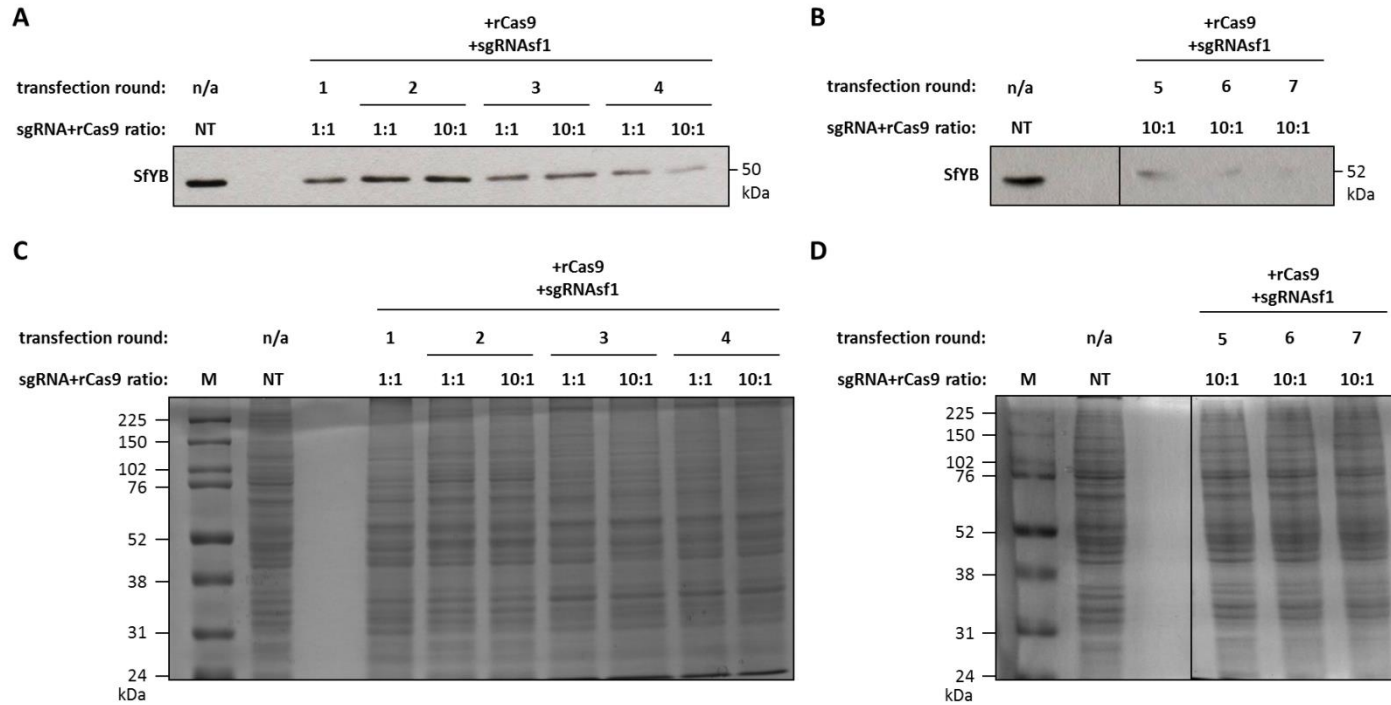


Figure 6.8 Screening multi-run transfected Sf9 cells with sgRNAsf1 for SFYB disruption. Western blot analysis using α YB1 (PA5-19453) of multi-run transfected Sf9 cells with 1:1 or 10:1 ratio of sgRNAsf1+rCas9 to CRISPRmax™ Reagent, for **(A)** four rounds transfections, demonstrating disruption of SFYB protein expression (partial knockout phenotype) by the fourth consecutive transfection of Sf9 cells. This was especially the case at 10:1 ratio of sgRNAsf1+rCas9 to CRISPRmax™ Reagent. And **(B)** complete disruption at the sixth and seventh round of transfections for 10:1 ratio, only. Of which observations were compared to non-transfected Sf9 cells as control for baseline expression of SFYB protein. Equal loading of protein samples was demonstrated by **(C and D)** Coomassie® Blue gel staining for the first four and final three transfection rounds, respectively. M, ECL™ Rainbow™ Marker – Full Range; n/a, not applicable; NT, non-transfected control; rCas9, recombinant Cas9.

Western blot analysis with antiserum raised against rSfYB(His)₁₀ (described in Chapter 5, section 5.3.5) showed a complete knockout of the SfYB protein expression from as early as the first round of transfection (1:1 ratio) of the multi-run transfection using sgRNAsf1 (Figs. 6.9A). This was with normalised protein loads demonstrated by Coomassie® Blue gel staining (Fig. 6.9B). However, the knockout phenotype was suggested transient, and the restoration of the SfYB phenotype occurred when transfected Sf9 cells were cultured for an extended period – by passage 15 and 25 post-transfection (Fig. 6.9A, see lanes p+15 and p+25, respectively).

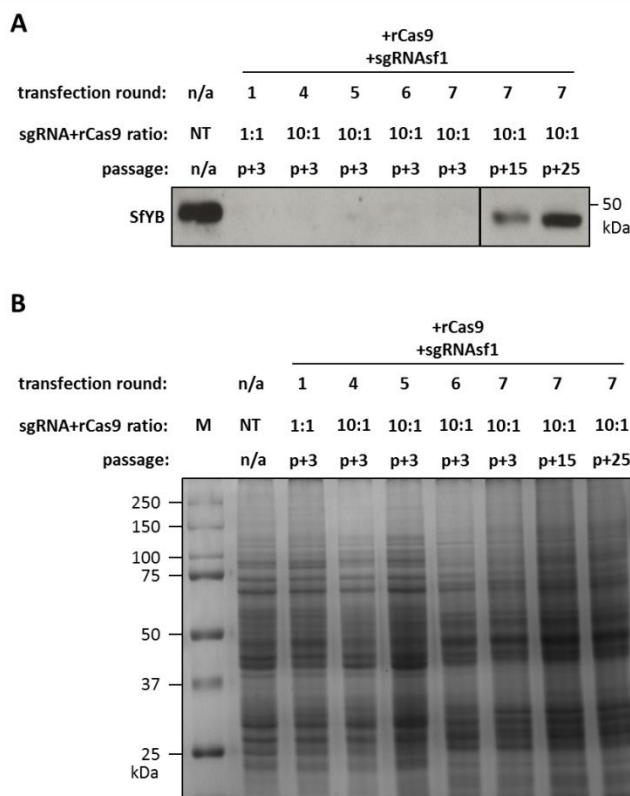


Figure 6.9 Confirmation of partial knockout and knockout phenotypes using SfYB-specific antiserum. *SfYB* gene disruption was confirmed with SfYB-specific antiserum, where **(A)** Western blotting analysis indicated complete knockout of SfYB expression as early as the first round of transfection and passage 3 post-transfection with sgRNAsf1 and recombinant Cas9 (1:1 ratio to CRISPRmax™ Reagent). Additionally, the eventual loss of the knockout phenotype was demonstrated by passage 15 and 25 post-transfection. Equal loading of protein was depicted by **(B)** Coomassie® Blue gel staining. M, Precision Plus Protein™ Dual Colour Standards; n/a, not applicable; NT, non-transfected control; p+3, passage 3 post-transfection; p+15, passage 15 post-transfection; p+25, passage 25 post-transfection; rCas9, recombinant Cas9.

Western blotting for Sf9 cells transfected with sgRNAsf2 initially indicated potential SfYB disruption after multi-run transfection for CRISPR/Cas9 genome editing, using α YB1 (Fig. 6.10A). This was in contrast to Western blott analysis using SfYB-specific antiserum, which revealed comparable SfYB protein expression between control Sf9 and transfected Sf9 cells, regardless of the increased ratio of sgRNAsf2 plus recombinant Cas9 to CRISPRmax™ (Fig.

6.10B), and normalised protein loading demonstrated by Coomassie® Blue gel staining (Fig. 6.10C). Sf9 transfection with sgRNAsf3 also displayed a similar disparity between α YB1 and SfYB-specific antiserum staining of blots for Western blot analysis (Figs. 6.11A-C). Therefore, it was concluded that targeted disruption of the *SFYB* gene was mainly achieved using sgRNAsf1 and recombinant Cas9 at 10:1 ratio to CRISPRmax™ Reagent. These cells were termed Sf9ex1.4-7/10X cells, reflective of the exon target, round of transfection, and transfection condition.

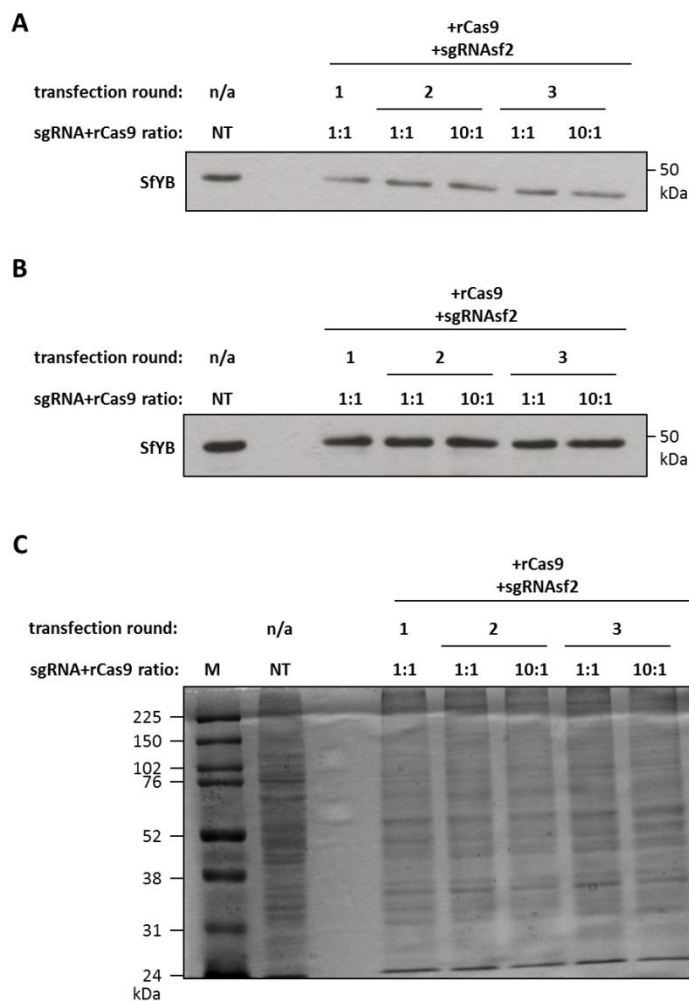


Figure 6.10 Screening for SfYB knockout cell lines transfected with sgRNAsf2 for *SFYB* disruption. A) Western blot analysis using α YB1 (PA5-19453) of multi-run transfected Sf9 cells (1:1 or 10:1 ratio of sgRNAsf2+rCas9 to CRISPRmax™ Reagent), indicating potential disruption of SfYB protein expression, especially when compared to non-transfected Sf9 control for baseline expression of SfYB. However, **(B)** staining with SfYB-specific antiserum instead suggests that SfYB disruption was not achieved and comparable SfYB protein expression was detected to that of non-transfected Sf9 control. Normalised protein loads, especially between transfected Sf9 cells, was determined by **(C)** Coomassie® Blue gel staining. M, ECL™ Rainbow™ Marker – Full Range; n/a, not applicable; NT, non-transfected; rCas9, recombinant Cas9.

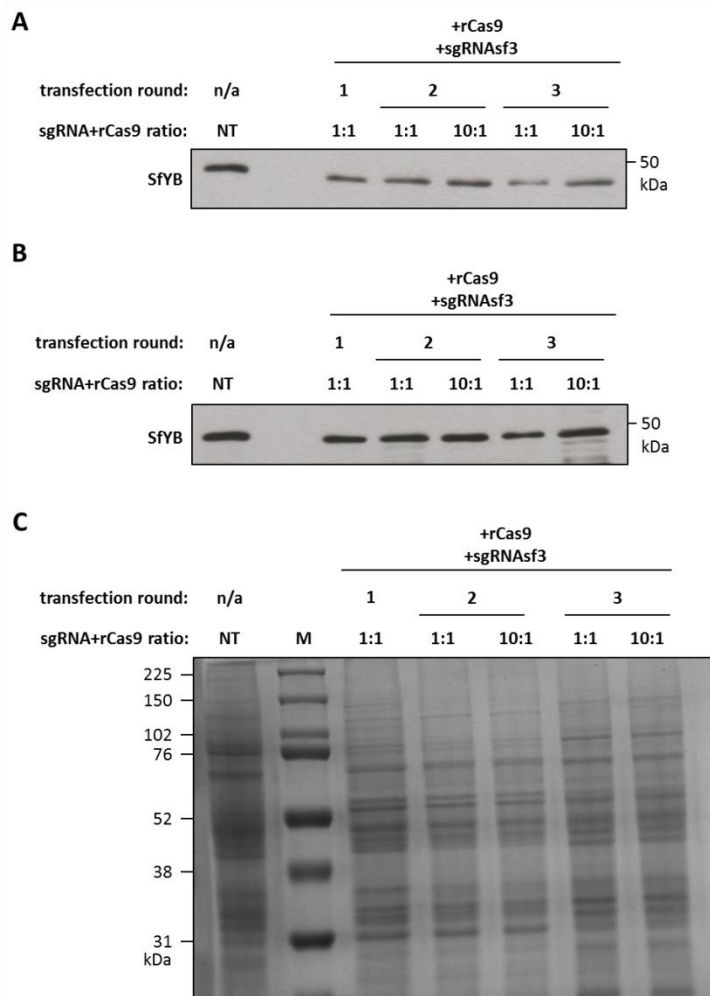


Figure 6.11 Screening for SfYB knockout cell lines transfected with sgRNAsf3 for *SFYB* disruption. A) Western blot analysis using α YB1 (PA5-19453) of multi-run transfected Sf9 cells (1:1 or 10:1 ratio of sgRNAsf3+rCas9 to CRISPRmax™ Reagent), indicating some potential disruption of SfYB protein expression, especially when compared to non-transfected Sf9 control for baseline expression of SfYB. However, **(B)** Western blot analysis using SfYB-specific antiserum suggested that SfYB disruption was not achieved using sgRNAsf3. This was especially the case were comparing relative protein expression to that of non-transfected Sf9 control, and **(C)** normalised loading of protein samples demonstrated by Coomassie® Blue gel staining. n/a, not applicable; NT, non-transfected; M, ECL™ Rainbow™ Marker – Full Range; rCas9, recombinant Cas9.

Sf9ex1.6/10X and Sf9ex1.7/10X cells were further analysed given their knockout phenotypes described by the limits of detection by Western blotting using SfYB-specific antiserum. Scale-up of the selected SfYB-disrupted Sf9 cell lines were associated with a notable lag in cell growth when compared to parental Sf9. This was especially notable when cells were expanded in suspension culture. In fact, a significant growth defect was observed by the analysis of growth curves and number of live cells/mL between parental Sf9 and selected Sf9ex1.6/10X or Sf9ex1.7/10X over a two-week timecourse (Fig. 6.12A, compare black growth curve with blue or red growth curves for parental Sf9 and Sf9ex1.6/10X and 1.7/10X, respectively). After two days in suspension culture the SfYB-disrupted cell lines exhibited a significant lag in cell growth

and delay into log-phase growth compared to Sf9 control, especially Sf9ex1.7/10X cell line (n = 3 each cell line counted; $P < 0.001$ compared to Sf9 control for all time points from day three onwards in suspension culture). Additionally, peak live cell density of $>1 \times 10^7$ cells/mL was not achieved by the SfYB-disrupted cell lines. The growth defect was also concluded to be independent of an increased rate in cell death as SfYB-disrupted cells showed improved cell viability over the two-week growth course compared to Sf9 control (Fig. 6.12B). Both Sf9 control and SfYB-disrupted cells showed comparable levels of viable cells (%) up to nine days in suspension. After which, SfYB-disrupted cell lines, particular Sf9ex1.7/10X, showed significantly improved cell viability (n = 3 each cell line counted; $P < 0.001$).

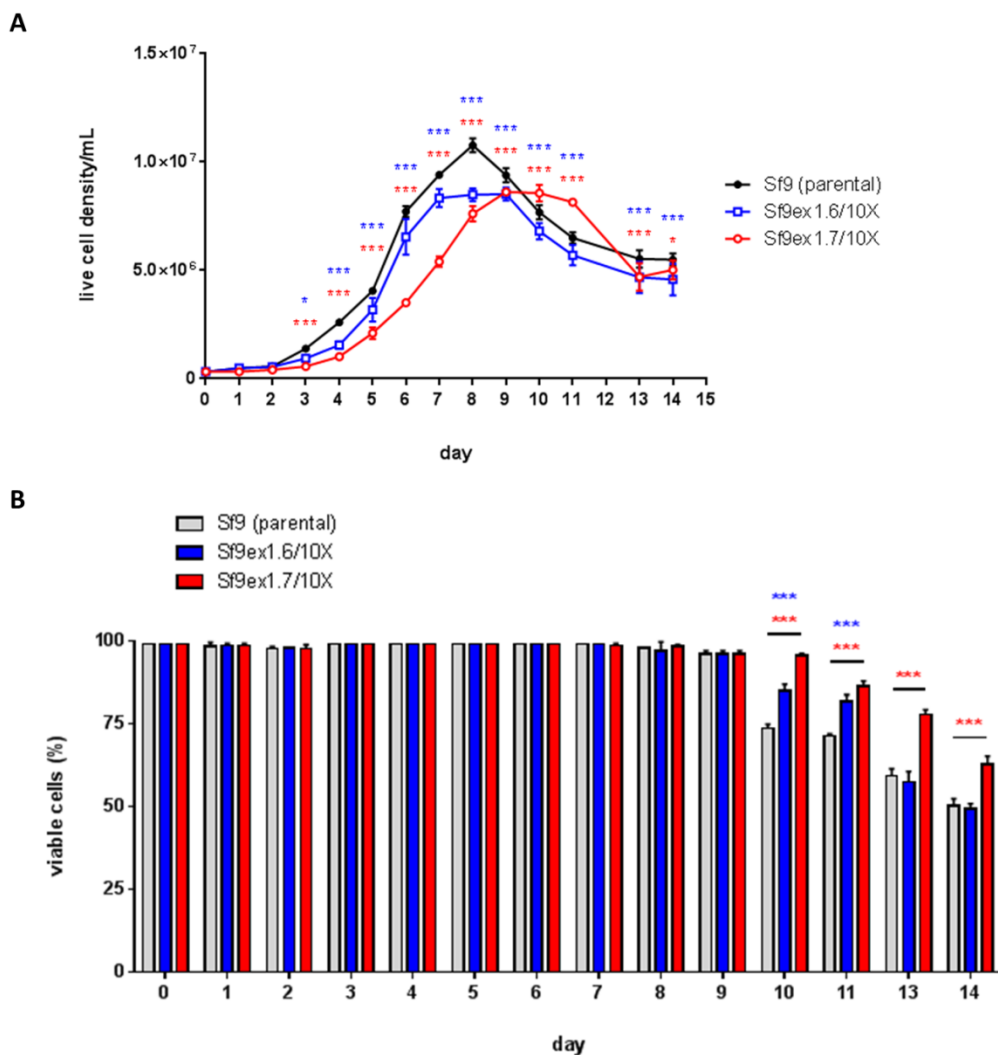


Figure 6.12 Growth defect of SfYB knockout cell line compared to Sf9. Parental Sf9 and selected Sf9 knockout cell lines (Sf9ex1.6-7/10X) were cultured in suspension, and cell growth kinetics was measured on a daily basis for up to 2 weeks. **A)** Significant growth defect was measured from day 3 onwards for Sf9ex1.6/10X (blue line and open square) or Sf9ex1.7/10X cells (red line and open circle) when compared to Sf9 control (black line and solid circle). The observed growth defect was confirmed to not be attributed by an increased rate in cell death, as **(B)** the cell viability (%) was improved for Sf9ex1.6 (blue solid bars) and 1.7/10X (red solid bars) compared to control Sf9 (grey solid bars), especially from day 10-14 in suspension culture. Figure legend continues next page.

Figure 6.12 Growth defect of SfYB knockout cell line compared to Sf9. Error bars reflect \pm SD from mean, $n = 3$, where averaged cell counts were from cells cultured in triplicate; *, $P < 0.05$, and ***, $P < 0.001$. Coloured * represents the calculated significance that was observed between control Sf9 and Sf9ex1.6/10X (blue) or 1.7/10X (red).

Characterisation of SfYB-disruption by CRISPR/Cas9 genome editing was further demonstrated by ICC and confocal laser scanning microscopy. Control Sf9 cells exhibited a diffuse cellular distribution of SfYB protein by confocal laser scanning microscopy (Fig. 6.13A, panels a-c). On the other hand, when examining different fields of view (FOV), SfYB-disruption was portrayed by Sf9ex1.7/10X cells stained with α SfYB(His)₁₀ with cells either exhibiting no SfYB protein staining (Fig. 6.13A, FOV 1 and 2, panels d-f and g-i, respectively), or significantly reduced SfYB expression (Fig. 6.13A, FOV 3, panels j-l, respectively) when compared to Sf9 control cells. Partial staining of SfYB expression in Sf9ex1.7/10X cells indicated that a heterozygous SfYB-disruption phenotype was achieved in transfection of Sf9 cells with sgRNAsf1 and recombinant Cas9. Negative staining controls showed background or negligible fluorescence exhibited by control Sf9 cells stained with primary antiserum or secondary antibodies only (Fig. 6.13A, panels m-o and p-r, respectively). The above findings were also confirmed by orthogonal section analysis (Fig. 6.13) to account for the z-plane of stained cells. SfYB predominantly was localised throughout the cytoplasm but detectable in the nuclear compartment as well, as per DAPI-staining (Fig. 6.13B). However, Sf9ex1.7/10X cells demonstrated no SfYB protein staining throughout the cell z-plane by orthogonal section analysis, indicating SfYB knockout (Fig. 6.13C).

In order to quantify the extent of SfYB-disruption, corrected total cell fluorescence (CTCF) was compared between Sf9 control as baseline reference and Sf9ex1.7/10X cells and Sf9 secondary antibody only staining control (Fig. 6.13D). Control Sf9 cells exhibited an average of 4255.37(\pm 1526.56) CTCF, which was significantly different to the average CTCF of 705.43(\pm 204.3) for Sf9 cells stained with only secondary antibody (Fig. 6.13D, compare black solid bar and grey solid bar; $P < 0.001$). However, the collectively analysed Sf9ex1.7/10X cells showed an average 843.28(\pm 690.15) CTCF, which was comparable to Sf9 cells stained with secondary antibody only (Fig. 6.13D, compare red solid bar with grey solid bar; $P > 0.05$). The 843.28(\pm 690.15) CTCF was also significantly different to the Sf9 baseline reference of 4255.37(\pm 1526.56) CTCF (Fig. 6.13D, compare black solid bar with red solid bar $P < 0.001$). Therefore, despite the heterozygous SfYB-disrupted phenotype characterised by ICC and confocal laser microscopy, a significant reduction in SfYB staining was overall quantified.

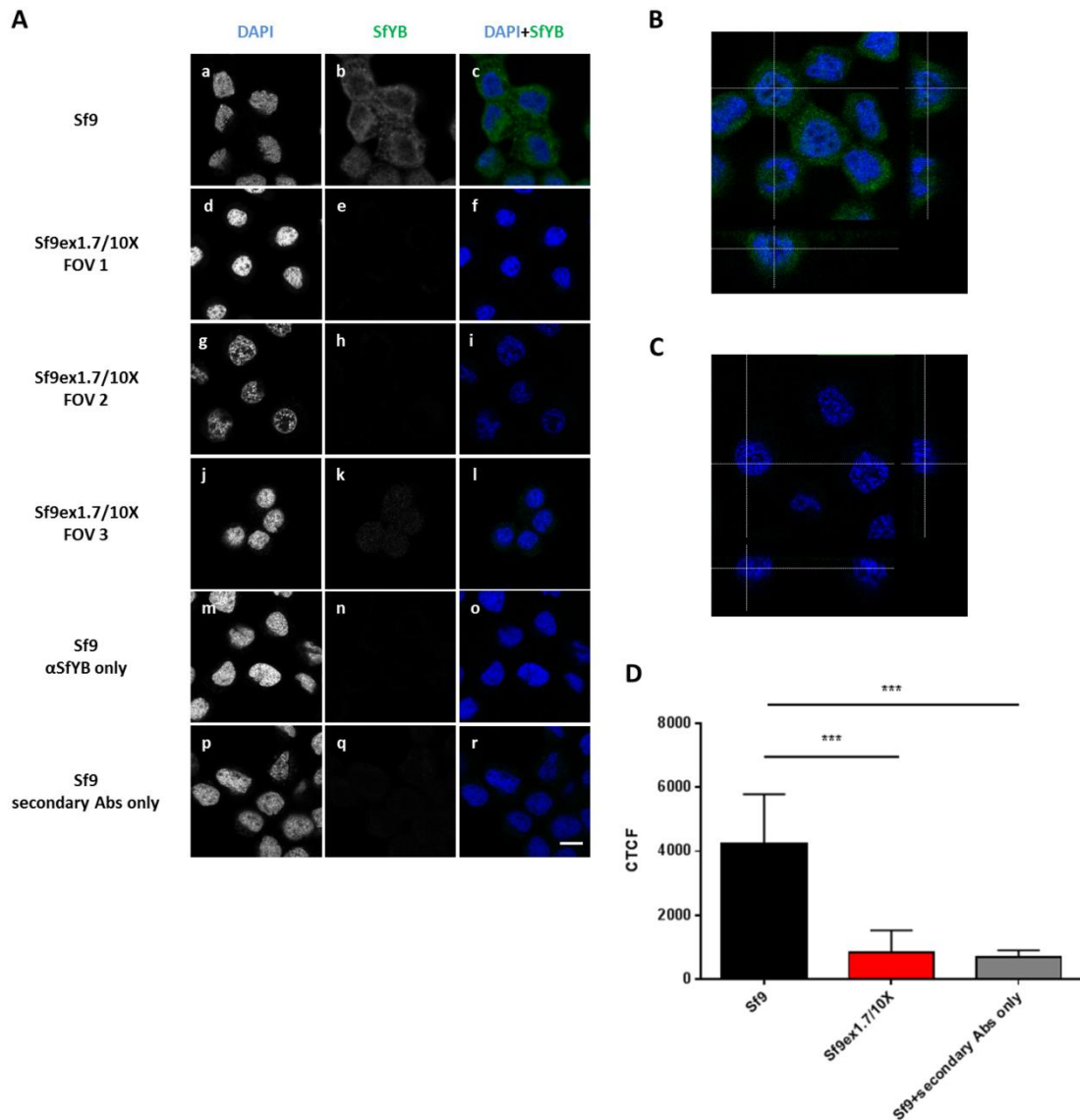


Figure 6.13 Characterising SfYB knockout cell lines using ICC and confocal laser scanning microscopy. The SfYB knockout phenotype was further verified by (A) ICC and confocal laser scanning microscopy of Sf9 and Sf9ex1.7/10X. Cells were stained with DAPI to differentiate the nuclei (a, d, g, j, m, and p), and SfYB-specific antiserum (b, e, h, k, n, and q). Panels (a-c) refer to control Sf9 cells and panels (d-l) refer to three fields of view of Sf9ex1.7/10X cells stained for SfYB, indicating that SfYB disruption was achieved. Negative staining controls (Sf9) with either primary α SfYB (m-o) or secondary antibody (p-r) exhibited no or background fluorescence. Representative scale bar = 20 μ m for all microscopy images. Orthogonal section analysis of (B) Sf9 and (C) Sf9ex1.7/10X for SfYB expression demonstrated that SfYB expression was diffuse throughout the cell when considering x-, y-, and z- planes, but no SfYB protein staining was detected in Sf9ex1.7/10X cells. (D) Graph presenting CTCF measurements between Sf9 control, Sf9ex1.7/10X, and Sf9 cells stained with only secondary antibody (negative staining control), indicating that Sf9ex1.7/10X exhibited fluorescence similar to the negative staining control. Error bars reflect \pm SD from mean, n = 30 measured cells for each cell line; except Sf9 stained with secondary antibody only control, n = 5; ***, $P < 0.001$ or $P < 0.0001$. Abs, antibody; CTCF, corrected total cell fluorescence; FOV, field of view.

Characterisation of Sf9ex1.7/10X cell line was also performed at the genomic level and compared to Sf9 control. Sequence chromatograms of the sgRNAsf1-specific and CRISPR/Cas9 targeted sequence of Sf9ex1.7/10X identified potential mutations within *SFYB*'s putatively mapped exon 1 (Fig. 6.14A and 6.14B). This was located approximately 21bp downstream of the targeted PAM sequence. However, the wildtype sequence was also recognised by Sanger sequencing (Fig. 6.14A, compare top sequence read [control Sf9] with bottom sequence read [Sf9ex1.7/10X]), and thus confirms the heterozygous *SFYB*-disrupted phenotype identified by ICC and confocal microscopy.

Furthermore, Surveyor mutations screening positively identified that CRISPR/Cas9-mediated mutations were present, and likely corresponded to the mutations identified by Sanger sequencing (Fig. 6.14C and D). Here, whilst non-template control PCR demonstrated no non-specific PCR amplification, the expected amplicon of 452bp was produced for Sf9 control and for each selected *SFYB*-disrupted cell line, including the 1:1 mix of genomic template derived from control Sf9 and Sf9ex1.7/10X (Fig. 6.14C). Nuclease S treatment of annealed homo- or hetero- duplex DNA molecules gave rise to digestion bands of approximately 298bp and 154bp, along with the initial 452bp band, by agarose gel electrophoresis (Fig. 6.14D). This was particularly observed for CRISPR/Cas9 transfected Sf9 cells that demonstrated *SFYB*-disruption by Western blotting, and absent for Sf9 control. In fact, the later established *SFYB*-disrupted cell lines showed more prominent digestion bands, and indicated that the degree of genome editing using sgRNAsf1 and recombinant Cas9 improved after each transfection set (Fig. 6.14D). It was also interpreted that a heterozygous genotype was inferred by Surveyor mutation screening assays for each *SFYB*-disrupted cell line, which did not require the input of control Sf9 genomic DNA to simulate a control heterozygous genotype (Fig. 6.14D).

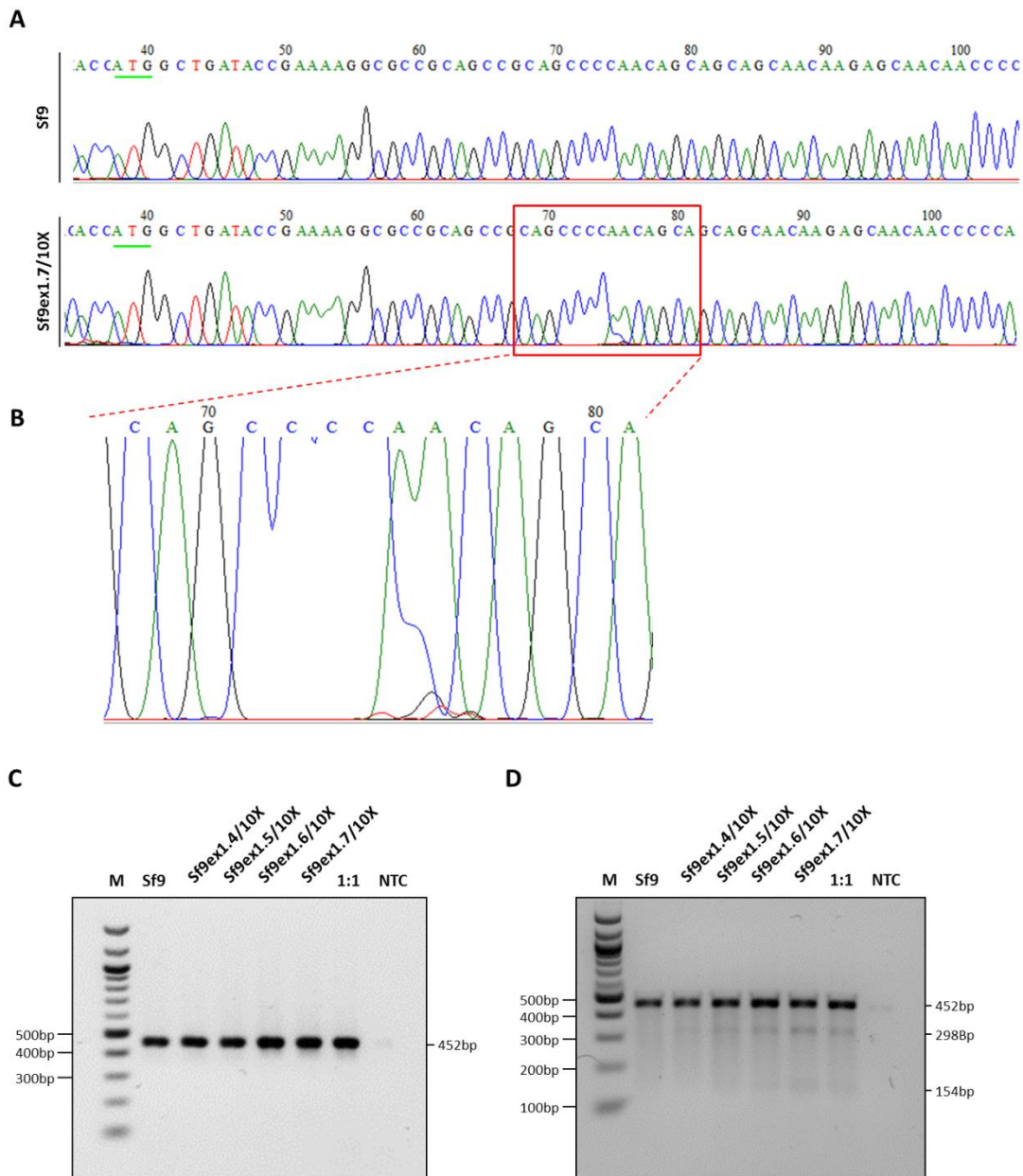


Figure 6.14 Genomic profiling of CRISPR/Cas9-mediated genome editing of the *SFYB* gene. Sf9ex1.7/10X CRISPR/Cas9-targeted DNA sequence was amplified and submitted for Sanger sequencing in parallel with Sf9 control. **A)** Chromatograms for Sf9 (top panel) and Sf9ex1.7/10X (bottom panel) sequencing results suggest potential mutations and a heterozygous genotype for Sf9ex1.7/10X sequence (red boxed), that is not present in Sf9 control. Green underlined sequence denotes translation start codon for *SFYB*. **B)** Closer examination of the Sf9ex1.7/10X chromatogram (red box expanded) implies some few mutations were recognised compared to Sf9 control, and confers a heterozygous genotype. Verification of the genotype was made by Surveyor™ Mutation Screening, where PCR amplification of the targeted gene sequence was confirmed as **(C)** approximately 452bp amplicons by 2% TBE agarose gel electrophoresis for Sf9 control, select Sf9ex 1.4/10X, 1.5/10X, 1.6/10X, and 1.7/10X cell lines, and 1:1 mix of Sf9 and Sf9ex1.7/10X genomic DNA as template (50ng total). PCR reactions were then treated with nuclease S and **(D)** digestion products corresponding to expected fragments (given the observed mutation) was detected (approximately 298bp and 154bp) for CRISPR/Cas9 treated cell lines, relative to control Sf9, by 2% TBE agarose gel electrophoresis. 1:1, 1:1 mix (50ng total) of genomic template from Sf9 control and Sf9ex.17/10X; M, 100bp DNA ladder; NTC, non-template control.

6.2.4 SfYB Y-Box protein homologue associations with AAV

The SfYB Y-Box protein homologue harbours the propensity for DNA-binding as was identified by electrophoretic mobility shift assays (EMSAs) using dsDNA and different quantities (μg) of purified rSfYB(His)₁₀ (Fig. 6.15). Binding to DNA plasmids that correspond to AAV-specific plasmids (pAAV2-hrGFP and pFBGR) and non-AAV-specific plasmids (pJET1.2-FLuc and pEx™-1) as controls demonstrated non-specific binding, and thus significant shifts in electrophoretic mobility by agarose gel electrophoresis, for the selection of plasmids tested (Fig. 6.15A). This was despite the quantity of rSfYB(His)₁₀ used, where as low as 0.1 μg of rSfYB(His)₁₀ was sufficient to entirely shift the mobility of 200ng plasmid migration with accumulation of nucleic acid in sample wells (see boxed) when compared to plasmid with no input of rSfYB(His)₁₀ as reference (Fig. 6.15A, compare plasmid lanes with input quantities of rSfYB(His)₁₀ protein to 0 μg input). Similarly, use of DNA probes that corresponded to AAV- (*ITR* and *rep*) or Adenovirus- (*E2A*) DNA sequences, and control DNA probes (left or right *ITR* flanking sequences) also exhibited electrophoretic shift in mobility after binding to rSfYB(His)₁₀ (Fig. 6.15B). Especially, electrophoretic mobility shift was evident using higher inputs of rSfYB(His)₁₀ – from 0.1 μg input upwards (Fig. 6.15B, see EMS annotation and red boxed). Ultimately, the EMSAs performed indicated that SfYB protein is capable of binding to dsDNA species; however specificity for AAV- or Adenovirus- associated sequences was not entirely observed.

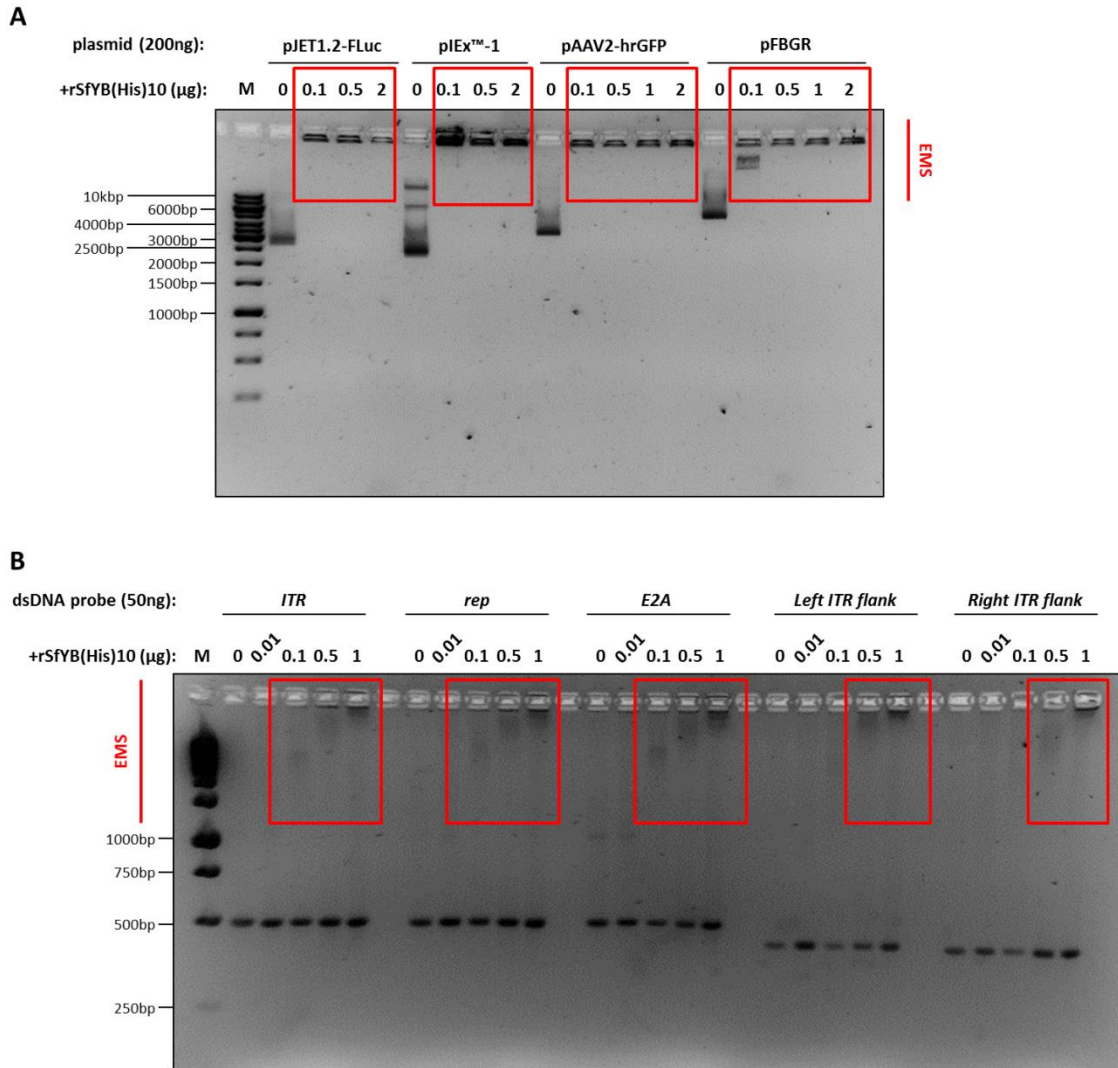


Figure 6.15 SfYB binding to DNA demonstrates shift in electrophoretic mobility by agarose gel electrophoresis. Recombinant SfYB protein showed DNA-binding propensity, with 200ng of plasmid DNA (pJET1.2-FLuc, pIExTM-1, pAAV2-hrGFP, or pFBGR) when mixed with different quantities of rSfYB(His)₁₀ in Cutsmart buffer. Plasmids demonstrated **(A)** shifts in electrophoretic mobility through a 2% TBE agarose gel for all input quantities of rSfYB(His)₁₀ (see EMS annotation and red boxes). Likewise, AAV- or AdV- associated dsDNA probes (50ng each of *ITR*, *rep*, *E2A*, and control left of right *ITR* flanking dsDNA probes) all showed **(B)** shifts in electrophoretic mobility through a 2% TBE agarose gel for 0.1, 0.5, and 1μg input of rSfYB(His)₁₀ (see EMS annotated and red boxes). EMS, electrophoretic mobility shift; L *ITR* flank; left *ITR* flanking control DNA probe; M, GeneRuler 1kb DNA Ladder; R *ITR* flank; right *ITR* flanking control DNA probe.

We also examined co-localisation of SfYB protein with AAV2 Rep, Cap, and particle by ICC and confocal laser scanning microscopy (Fig. 6.16). Whilst SfYB protein expression was positively detected throughout infected Sf9 control cells (Fig. 6.16, panels a-d, i-l, and q-t), and SfYB-disruption was reaffirmed by ICC of infected Sf9ex1.7/10X cell lines (Fig. 6.16, panels e-h, m-p, and u-x). Additionally, it was noted that SfYB protein expression was not necessarily predominantly localised to the cytoplasm of infected Sf9 cells with perhaps some SfYB localising to the nuclei (see Fig. 6.13, panels a-d, i-l, and q-t). This was in some contrast to previous observations of SfYB locale in non-infected Sf9 cells. Nonetheless, AAV2 Rep protein was identified localised to the nuclei for both Sf9 control and Sf9ex1.7/10X, as distinguished by DAPI-staining of the nuclear compartment (Fig. 6.16, panels a-d and e-h, respectively). Whereas, AAV2 Cap protein staining was found localised to both the cytoplasm and nucleus of infected Sf9 control and Sf9ex1.7/10X (Fig. 6.16, panels i-l and m-p, respectively). Finally, rAAV2 particle was found localised to the nucleus of either infected Sf9 or Sf9ex1.7/10X cell lines (Fig. 6.16, panels q-t and u-x, respectively). It was also acknowledged that single antibody staining of infected Sf9 cells for rAAV2 vector production demonstrated no (staining with primary antibodies only; Fig. 6.16, panels a-d) or minimal background (staining with secondary antibodies only; Fig. 6.16, panels e-h) fluorescence. Ultimately, co-localisation of SfYB protein with AAV2 Rep, Cap, or intact particle was positively identified by ICC and confocal laser scanning microscopy, which potentially infers possible interactions between SfYB and AAV2 proteins.

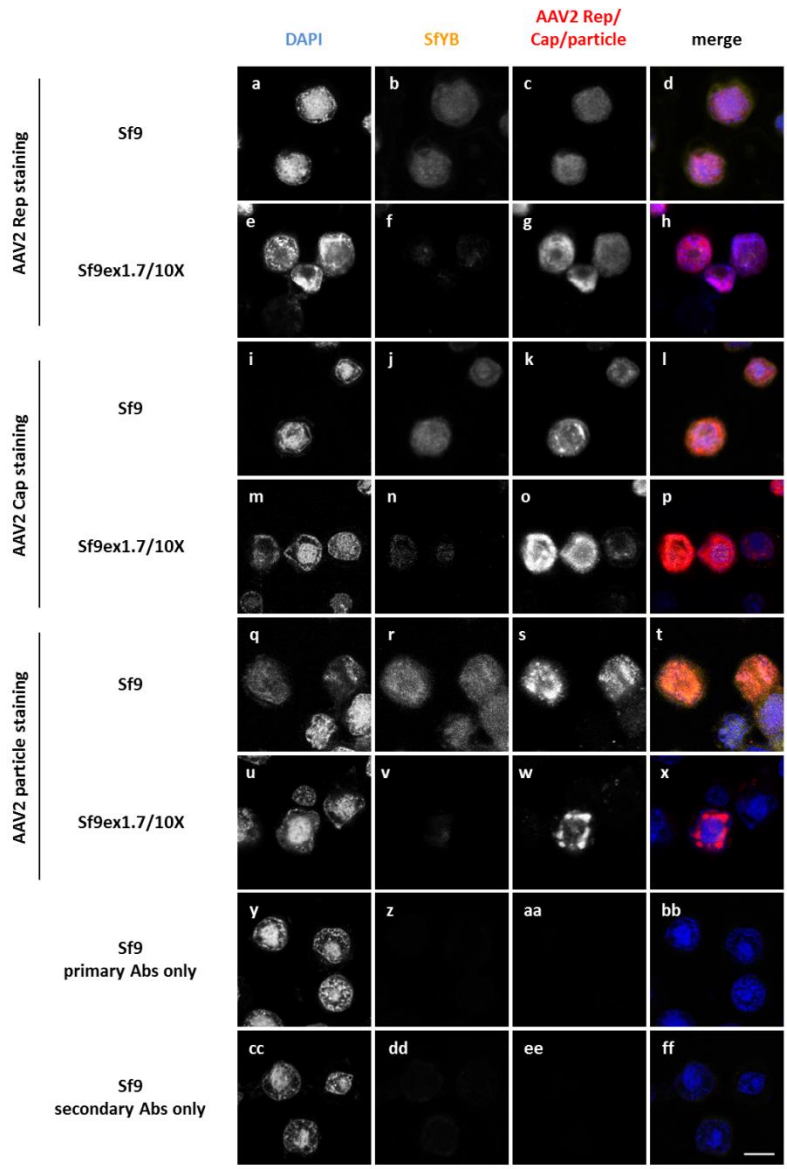


Figure 6.16 ICC and confocal laser scanning microscopy of SfYB and AAV2 Rep, Cap, and particle localisations. Sf9 and selected Sf9ex1.7/10X were infected for rAAV2GFP vector production, and cells were fixed, permeabilised and stained for SfYB and AAV2 proteins. DAPI staining served to distinguish the nuclear compartment. Panels (a-d) and (e-h) refer to infected Sf9 and Sf9ex1.7/10X (respectively) stained for SfYB and AAV2 Rep. Panels (i-l) and (m-p) refer to Sf9 and Sf9ex1.7/10X cells (respectively) stained for SfYB and AAV2 Cap proteins. Finally, panels (q-t) and (u-x) refer to Sf9 and Sf9ex1.7/10X cells (respectively) stained for SfYB and AAV2 intact particle. Single antibody staining controls of infected Sf9 cells with primary (y-bb) or secondary (cc-ff) antibodies demonstrated no to minimal background fluorescence. Representative scale bar = 10µm for all panels.

6.2.5 The effect of SfYB disruption on rAAV vector production

We next explored whether or not SfYB expression or disruption affected baculovirus infection when compared to Sf9 control. Therefore, the infection efficiency of BacITRGFP was examined in context of rAAV -2 or -8 vector production by flow cytometry (Fig. 6.17). We demonstrated that comparable BacITRGFP infection efficiencies (>95%) were presented by SfYB-disrupted cells and control Sf9 for rAAV -2 or -8 vector production (Fig. 6.17; compare green and blue lines, and red and pink lines, respectively). Additionally, it was considered that the infection of BacSfYB(His)₁₀ and expression of exogenous SfYB (rSfYB(His)₁₀) did not significantly hinder BacITRGFP infection. More specifically, an infection rate of 89.4% was demonstrated for rAAV2 vector production co-infected with BacSfYB(His)₁₀ compared to 95% for Sf9 cells co-infected with control BacMers-M (Fig. 6.17). The above was relative to mock infection control (Sf9 cells infected with BacMers-M only; black line), which confidently resulted in no/low proportion of cells with detectable fluorescence (0.35%). Therefore, SfYB-disruption or its over expression did not affect baculovirus infection.

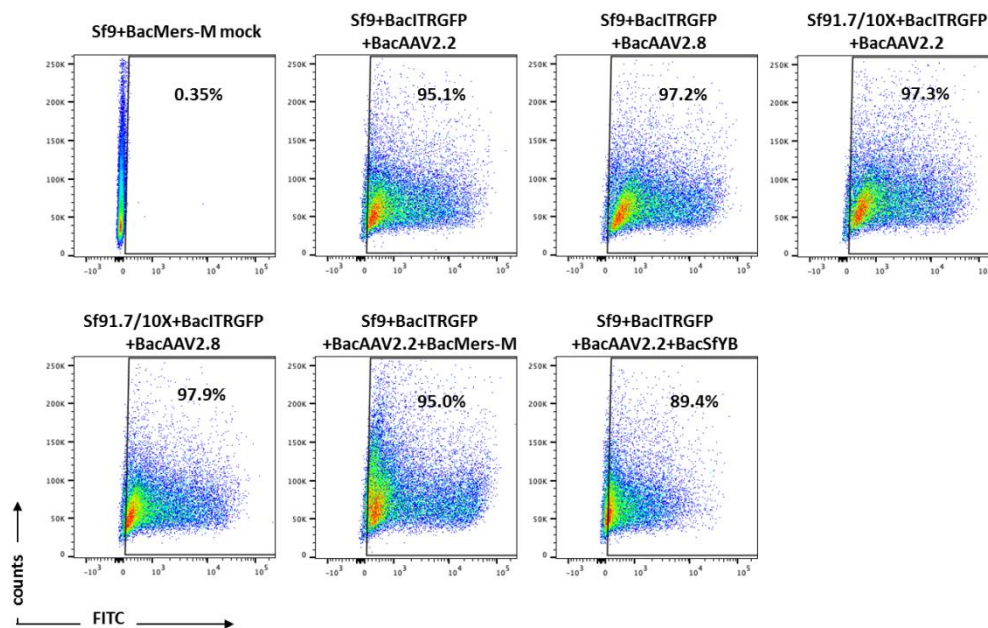


Figure 6.17 SfYB disruption or exogenous expression does not affect the efficiency of baculovirus infection. The infection of Sf9 and Sf9ex1.7/10X cells with MOI 3 baculoviruses and analysed by flow cytometry and scatter plots, 72h post-infection. BacMers-M infection of Sf9 functioned as a mock infection control for transduction rates. Flow cytometry analysis suggested that BacITRGFP infection efficiency was comparable between parental Sf9 and Sf9-disrupted cell lines for rAAV -2 or -8 vector production (rates of >95% was recorded). Additionally, the infection rate of 89.4% and 95% was considered relatively indifferent with the co-infection of BacSfYB(His)₁₀ or BacMers-M, respectively, indicating that exogenous SfYB expression did not compromise the transduction rate of Sf9 cells for BacITRGFP. The % of fluorescent cells is displayed within scatter plots.

Given we previously reported that the YB1 knockout phenotype of 293T compromised the cell line's resistance to chloroquine reagent (see Chapter 3 and 4), we examined whether SfYB disruption made the cell line susceptible to cell cytotoxicity by baculovirus infection for rAAV vector production. Non-infected Sf9 and Sf9ex1.7/10X cells demonstrated low but comparable cytotoxicity profiles, with %cytotoxicity of $1.05 \pm 0.05\%$ and $1.12 \pm 0.1\%$ for Sf9 (white solid bar) and Sf9ex1.7/10X (black solid bar), respectively (Fig. 6.18A, $n = 3$, $P > 0.05$). This correlated and was in agreement with viable cell counts at day 3 observed previously (Fig. 6.12B). Similarly, comparable %cytotoxicity was calculated between Sf9 ($13 \pm 0.27\%$, white solid bar) and Sf9ex1.7/10X ($11.8 \pm 0.85\%$, black solid bar) cells after baculovirus infection for rAAV vector production (Fig. 6.18A; $n = 3$, $P > 0.05$). It was further appreciated that the infection efficiency of BacITRGFP was considered comparable between both cell lines ($>93\%$), as was determined by flow cytometry analysis (Fig. 6.18B). This suggested that an incomparable infection rate did not confound the cytotoxicity profiles calculated above. Therefore, LDH cytotoxicity assays identified that SfYB-disruption did not compromise the cell lines' susceptibility to cell cytotoxicity after baculovirus infection.

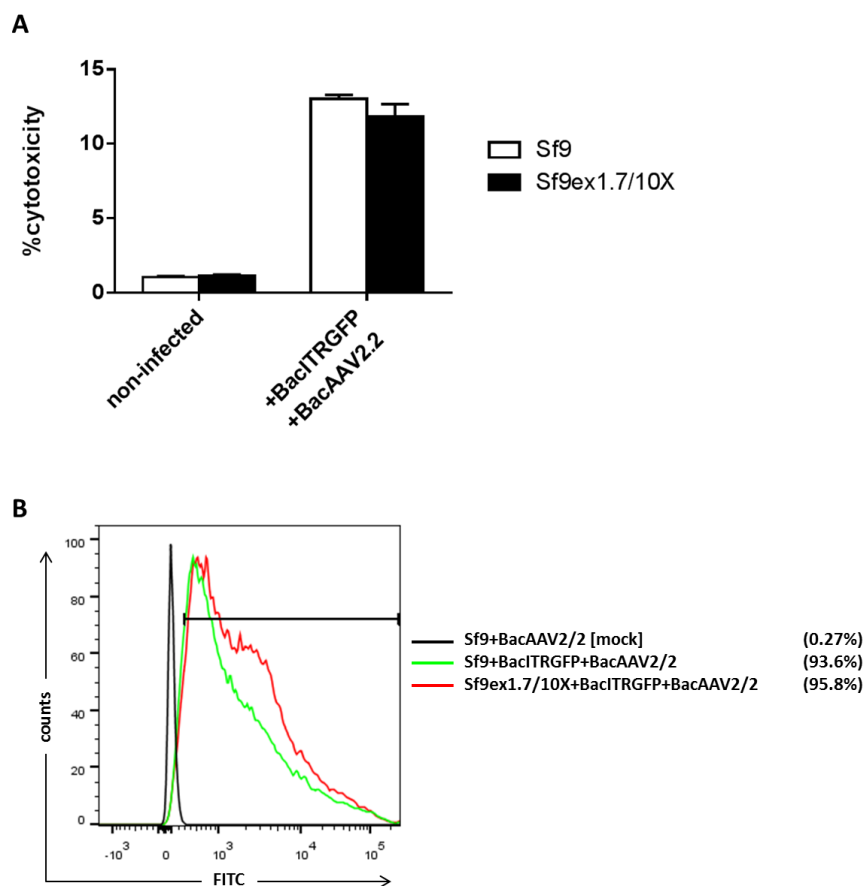


Figure 6.18 LDH cytotoxicity assays for baculovirus infection of Sf9 and SfYB knockout cell lines. Figure legend continues next page.

Figure 6.18 LDH cytotoxicity assays for baculovirus infection of Sf9 and SfYB knockout cell lines. Parental Sf9 and Sf9ex1.7/10X cells were infected with recombinant baculoviruses for rAAV2 vector production, and **(A)** %cytotoxicity measured 72h post-infection using LDH cytotoxicity assays. This inferred that comparable %cytotoxicity was measured for infected Sf9 and Sf9ex1.7/10X cell lines. Non-infected cells functioned as negative controls for baseline cytotoxicity profiles, and were harvested in parallel. Error bars reflect \pm SD from mean, $n = 3$; $P > 0.05$. Infection efficiency of BacITRGFP infection was also measured by **(B)** flow cytometry, with analysis indicating that comparable infections for rAAV2 vector production was achieved between parental Sf9 and Sf9ex1.7/10X cell lines (>93%), and compared to Sf9 mock infected negative control.

An initial indication that rAAV vector production may be influenced by SfYB-disruption was inferred by gene transfer assays of 293T with crude rAAVGFP vector lysates (Fig. 6.19). Firstly, comparable baculovirus infection efficiencies of approximately 80% between rAAV -2 or -8 vector productions were determined by flow cytometry analysis (Fig. 6.19A). This was further compared against mock infected Sf9 (0.34%) and Sf9 infected with only BacITRGFP and BacMers-M (88.6%) as negative and positive controls for baculovirus infection, respectively (Fig. 6.19A).

Therefore, the gene transfer rates for rAAV -2 or -8 vector particles encoding GFP transgene, derived from control Sf9 (SfAAV vectors) or Sf9ex1.7/10X (Sf1.7AAV vectors) producers, were analysed. Firstly, significant heat inactivation of baculovirus particles for gene transfer assays was indicated by comparing 293T infections with non-heat inactivated and heat inactivated crude lysates derived from control Sf9 cells infected with BacITRGFP+BacMers-M only (Fig. 6.19B and C). A drastic reduction in transduction rates from 62% to 1.16% for non-heat inactivated and heat inactivated control lysates, respectively, was confidently observed (Fig. 6.19B and C). Therefore, with BacITRGFP heat inactivated from sample crude AAV lysates, a notable improvement in gene transfer was achieved after 293T infections with Sf1.7AAV vector when compared to SfAAV for most dilutions tested (Fig. 6.19A and B, respectively for crude AAV2 and AAV8 gene transfers). For instance, infections of 293T with heat-inactivated crude Sf1.7AAV2 was able to outperform SfAAV2 infections at the respective dilutions – neat, 1/5, and 1/25. Infection with neat, heat-inactivated Sf1.7AAV2 resulted in 58% gene transfer compared to 33.6% for SfAAV2 at the equivalent dilution (Fig. 6.19B). However, the effect on gene transfer was less apparent between rAAV8 vectors derived from parental Sf9 or Sfex1.7/10X cell lines (Fig. 6.19C). In fact, only infections with neat Sf1.7AAV8 vector (heat inactivated) exhibited improved gene transfer rates (72.2%) compared to SfAAV8 (66.4%) (Fig. 6.19C).

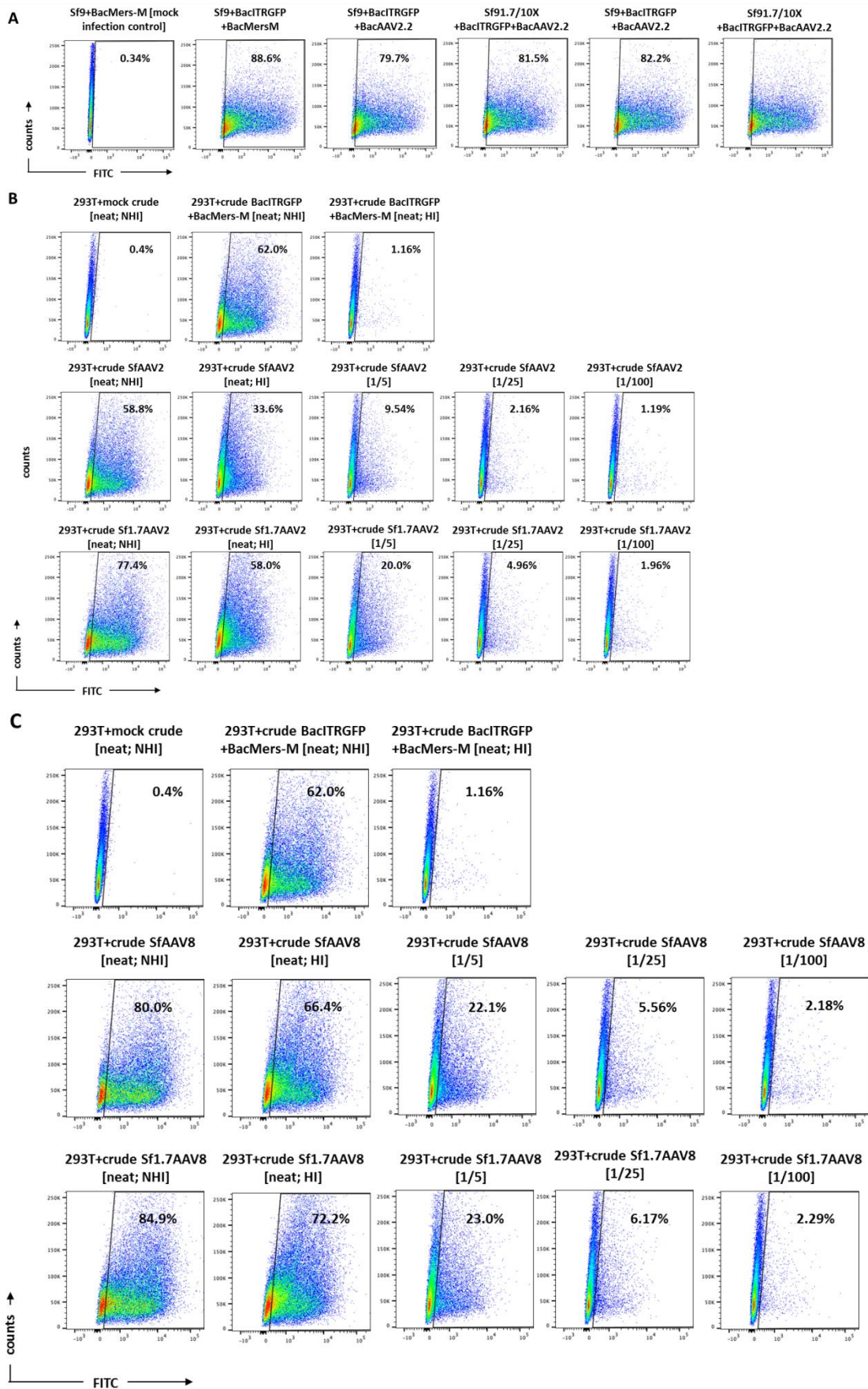


Figure 6.19 Gene transfer assays of 293T with crude rAAV -2 or -8 vectors. Figure legend continues next page.

Figure 6.19 Gene transfer assays of 293T with crude rAAV -2 or -8 vectors. The infection efficiency of BacITRGFP was comparable as determined by **(A)** flow cytometry and scatter plots of Sf9 and Sf9ex1.7/10X cells infected with baculoviruses for rAAV2 and rAAV8 vector production (approximately 80%). Sf9 cells infected with only BacMers-M functioned as mock infection control (0.34%), and cells infected with only BacITRGFP+BacMers-M functioned as positive control for baculovirus infection (88.6%). Crude lysates were then harvested by freeze-thaws, heat inactivated, and subsequently applied as a titration to 293T cells for gene transfer assays for **(B)** rAAV2 or **(C)** rAAV8 vectors. Flow cytometry analysis indicated that modestly improved gene transfer with rAAV vector particles derived from Sf9ex1.7/10X cells (Sf1.7AAV), particularly Sf1.7AAV2, was observed when compared to rAAV -2 or -8 vectors from Sf9 (SfAAV -2 or -8, respectively). Crude lysates from mock infected Sf9 served as mock crude infection control, and lysates derived from Sf9 cells infected with only BacITRGFP+BacMers-M functioned to test the efficiency of BacITRGFP heat inactivation. The baculovirus infection or gene transfer rates are displayed as % in scatter plots. HI, heat inactivated; NHI, non-heat inactivated; SfAAV; rAAV vector derived from parental Sf9; Sf1.7AAV, rAAV vector derived from Sf9ex1.7/10X.

Having determined that the SfYB-disrupted phenotype in Sf9 did not compromise the cell line's capacity for infection with baculovirus(es), or cell cytotoxicity post-infection, the expression levels of AAV2 Rep and AAV Cap (Vp1-3) proteins between cell lines was examined. Western blot analysis suggested that AAV2 Rep and AAV2 or AAV8 Vp1-3 protein expression was relatively comparable between Sf9ex1.7/10X and Sf9 control (Fig. 6.20A, top and bottom panels, respectively). This was despite the fact that Sf9ex1.7/10X cells demonstrated a consistent SfYB knockout phenotype (Fig. 6.20A, bottom panel), and equal protein loads were demonstrated by Coomassie® Blue gel staining between either baculovirus-infected cell lines for rAAV -2 or -8 vector production (Fig. 6.20B). Including, infection efficiencies of baculoviruses (BacITRGFP) was considered comparable, recording between 80-82% for control Sf9 and Sf9ex1.7/10X cell lines by flow cytometry (Fig. 6.20C). Therefore, it was implied that the expression of AAV Rep or Cap proteins was relatively unaffected by the disruption of SfYB expression in Sf9 cells.

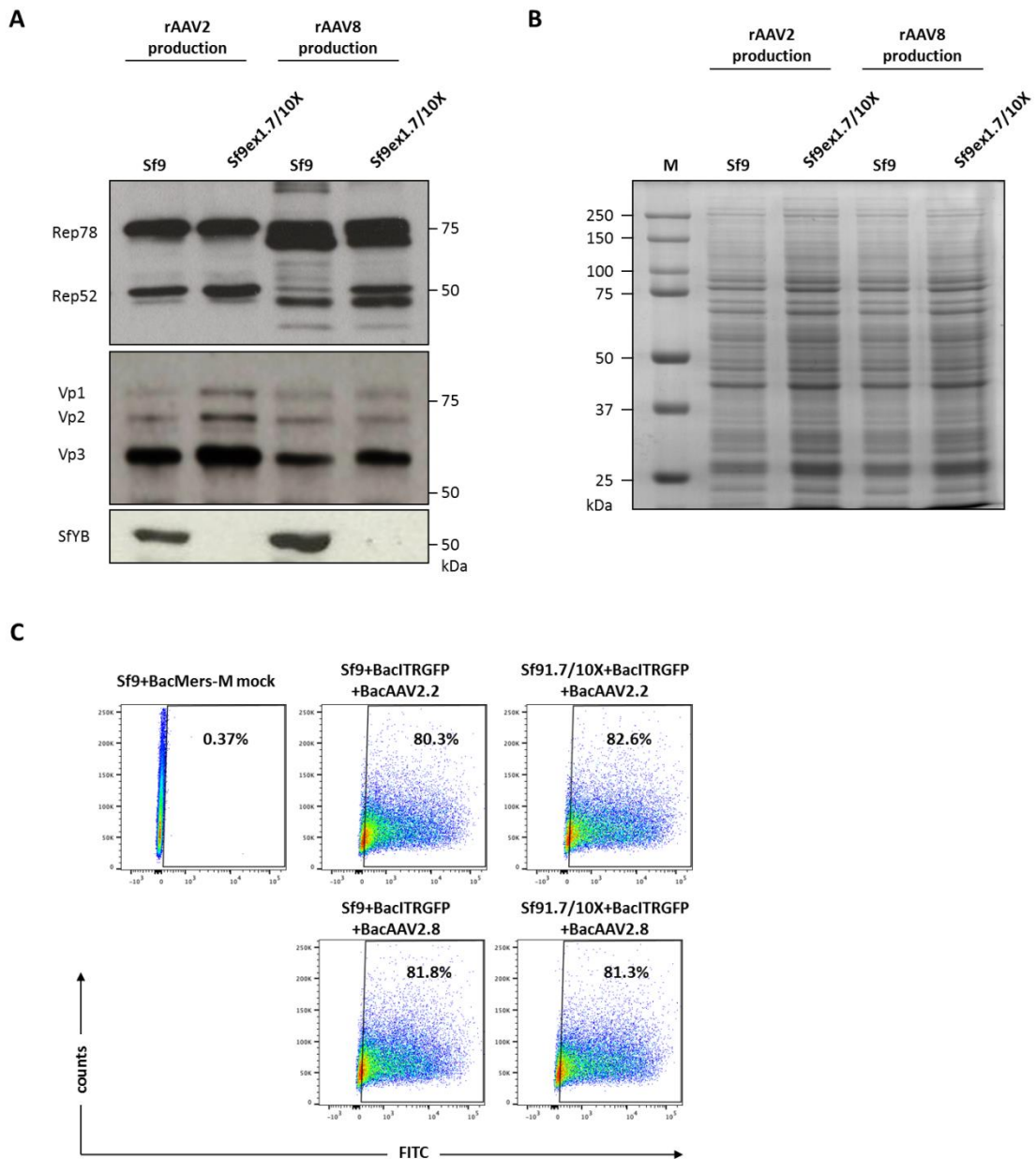


Figure 6.20 SfYB disruption did not influence AAV Rep or Cap (Vp1-3) protein expression. **A**) Western blot analysis revealed that the Rep -78 or -52 protein (top panel) and AAV2 and AAV8 Vp1-3 (middle panel) proteins were not influenced by SfYB disruption. Therefore AAV Rep and Cap expression in Sf9ex1.7/10X cells remained comparable to that of Sf9 control producer cells. SfYB disruption by CRISPR/Cas9 was demonstrated also using SfYB-specific antiserum (bottom panel). **B**) Equal loading of protein lysates was inferred by Coomassie® Blue gel staining, and **(C)** baculovirus (BacITRGFP) infection rates were considered comparable between Sf9 and Sf9ex1.7/10X rAAV vector productions (recording between 80-82%, compared to mock infected control at 0.37%). The % of infected cells is displayed in parentheses. M, Precision Plus Protein™ Dual Colour Standards.

Lastly, rAAV -2 or -8 vector genome titres from Sf9ex1.7/10X cells was calculated by qPCR, and compared relatively to control Sf9 as the baseline producer cell line. We first ensured that the efficiency of co-infection with baculoviruses (BacITRGFP) was comparable between producer cell lines for rAAV2 (Fig. 6.21A; compare green and red lines) and rAAV8 (Fig. 6.21A; compare

blue and pink lines) vector productions. Following which, qPCR analysis of vector genome titres indicated that despite SfYB disruption, the relative titres were considered comparable and not statistically significant to titres achieved by Sf9 control cells (Figs. 6.21B and C; $n = 3$ each, $P > 0.05$). Relative rAAV2 genome titres showed an average 2.56 ± 1.85 -fold more rAAV2 vector genome titres generated by Sf9ex1.7/10X to that of Sf9 control, but was deemed statistically insignificant (Fig. 6.21B; $n = 3$, $P > 0.05$). Similarly, an average 1.82 ± 0.64 -fold more rAAV8 vector genome titres was produced from Sf9ex1.7/10X to that of Sf9 control, and was also considered statistically insignificant (Fig. 6.21C; $n = 3$, $P > 0.05$). Batch-to-batch variation of physical rAAV -2 or -8 vector genome titres was observed, such that as much as 5.45×10^8 rAAV2 vector genomes/mL was measured for Sf9 compared to 3.67×10^8 vector genomes/mL for Sf9ex1.7/10X (batch 1), corresponding to 1.48-fold less (Fig. 6.21D). Or as low as 3.2×10^8 rAAV2 vector genomes/mL from Sf9 control compared to 8.4×10^7 vector genomes/mL from Sf9ex1.7/10X cells (batch 2), corresponding to 2.62-fold more vector genomes measured from the Sf9ex1.7/10X cell line (Fig. 6.21D). Likewise, as much as 1.02×10^9 rAAV8 vector genomes/mL from Sf9 compared to 1.57×10^9 vector genomes/mL from Sf9ex1.7/10X (batch 3), corresponded to 1.54-fold more rAAV8 (Fig. 6.21E). Or as low as 1.82×10^8 vector genomes/mL for Sf9 compared to 4.46×10^8 vector genome/mL from Sf9ex1.7/10X cells (batch 1), corresponding to 2.45-fold more rAAV2 vector genomes measured from the Sf9ex1.7/10X cell line (Fig. 6.21D). The above quantification of genome titres was confidently appreciated given that the purification quality of rAAV -2 or -8 vector preps from Sf9 and Sf9ex1.7/10X producer cells was demonstrated by silver staining, and distinguishable in purified samples compared to crude rAAV2 lysate control lane (Fig. 6.21F).

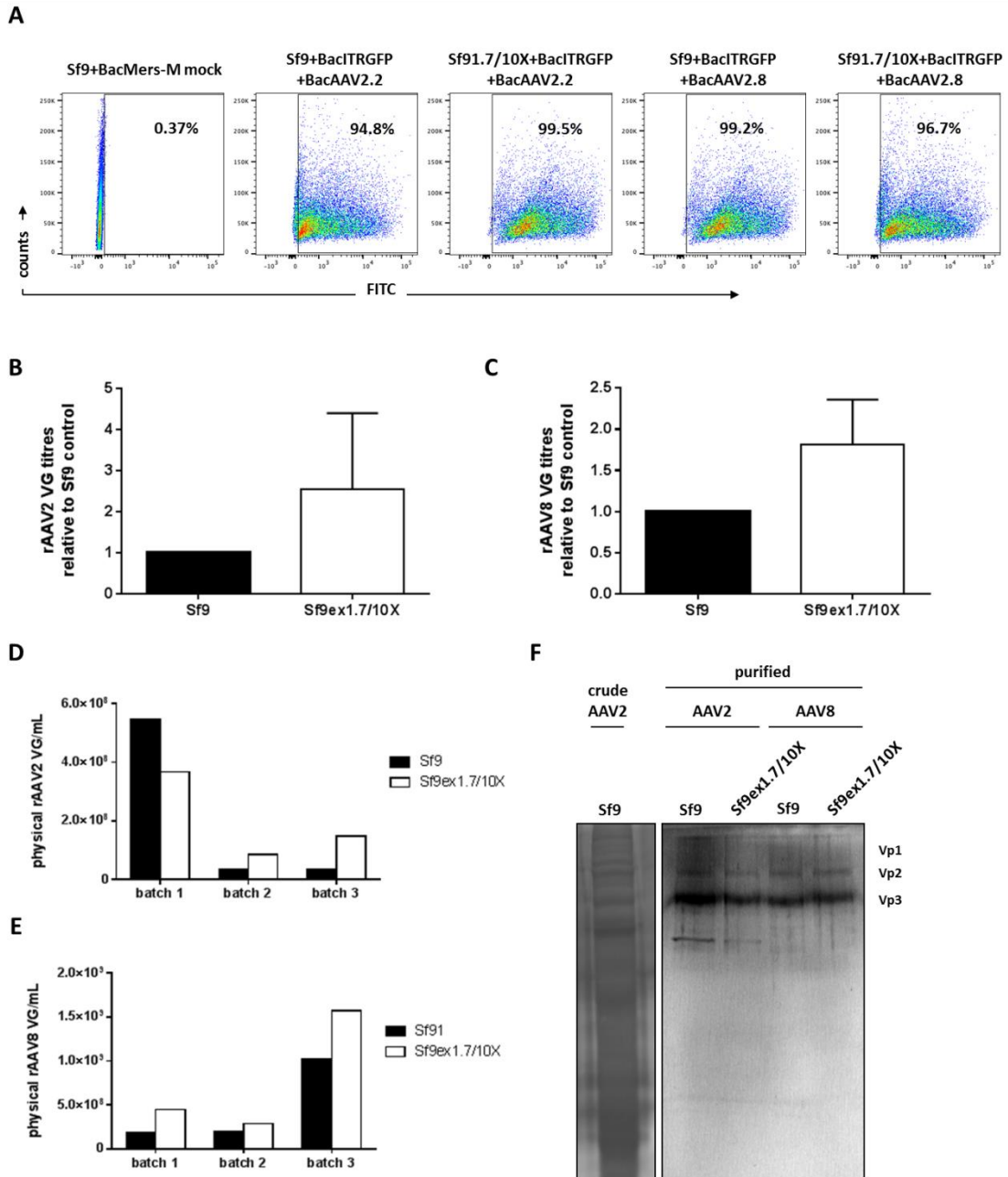


Figure 6.21 Comparable rAAV VG titres were detected between parental Sf9 and Sf9ex1.7/10X cell lines. The infection efficiency of BacITRGFP was determined by **(A)** flow cytometry, and considered comparable between control Sf9 and Sf9ex1.7/10X cell lines for rAAV -2 or -8 vector production, reporting at >94% of cells positively infected compared to mock infection control (0.37%). The % of fluorescent cells is displayed in parentheses. Recombinant AAV vectors were then harvested, purified, and subsequently quantified by qPCR **(B)** for relative rAAV2 and **(C)** rAAV8 VG titres, using Sf9 as the baseline control producer cell line. A statistically insignificant difference in relative VG titres for Sf9ex1.7/10X cells to that of Sf9 control was calculated by qPCR, indicating that SfYB-disruption did not confer an enhanced rAAV vector producer cell line ($P > 0.05$). Error bars reflect \pm SD from mean, $n = 3$. Individual physical **(D)** rAAV2 and **(E)** rAAV8 vector genomes/mL, as quantified by qPCR were plotted for each batch (1-3), and compared between control Sf9 and Sf9ex1.7/10X producer cells. **(F)** Silver staining of purified rAAV -2 or -8 vectors from Sf9 or Sf9ex1.7/10X demonstrated well purified rAAV vector preps with capsid proteins positively stained, and compared to crude rAAV2 harvest from Sf9 control. VG, vector genome.

6.3.0 Chapter summary

The present study demonstrated that genome editing of the Sf9 cell line was feasible, despite the restriction in promoter usage and limited reference genome data available. Regardless of this, *in silico* identification of the Y-Box protein homologue endogenous to *S. frugiperda* permitted the design of gRNAs for subsequent genome editing of the Sf9 cell line by CRISPR/Cas9 technology. However, the transient nature of the SfYB disruption seemed to be a consequence of the reduced growth kinetics when compared to parental Sf9, and being unable to establish single cell clones. In fact, the slower growing genome edited Sf9 cells were eventually outgrown by parental Sf9 cells in bulk CRISPR-transfected Sf9 cells for SfYB knockout, or reverted back to wildtype, eventuating in SfYB-expression back to wildtype levels. The benefit of single cell cloning CRISPR/Cas9 genome-edited cells was demonstrated in Chapter 3, and suggested such practice helped isolate homozygous YB1 knockout clones that were not in competition of parental 293T cells, or could not revert back to wildtype given the homozygous mutation profile(s). Although, this was in spite of the fact that Sf9 cells were originally clonal derivatives of the Sf21 cell line (Vaughn *et al.*, 1977). Nonetheless, SfYB disruption did not significantly exhibit an enhanced rAAV vector producer cell line quality. Recombinant rAAV vector genome titres and AAV Rep and Vp1-3 protein expression were all considered comparable to that of parental Sf9 as the control producer cell line.

Chapter 7: Discussion and Future Directions

7.1 Introduction

Recombinant Adeno-associated viral (rAAV) vector technology has grown to become one of the most promising tools for gene therapy applications. In fact, the successful implementation and licencing of gene therapy products for clinical and therapeutic administration has been demonstrated for rAAV. Key examples are rAAV products in the form of Glybera (Carpentier *et al.*, 2012) and Luxturna (Bennett *et al.*, 2016) for the treatment of hereditary lipoprotein lipase deficiency and Leber's congenital amaurosis, respectively. However, the demand for sufficient quantities of rAAV vector is constrained by current production systems. The importance of understanding host-AAV interactions is becoming more prevalent, with the importance of particular cell-intrinsic factors to AAV biology and processing slowly being revealed (Mitchell *et al.*, 2014; Satkunanathan *et al.*, 2014; Holscher *et al.*, 2015; Schreiber *et al.*, 2015).

The production of rAAV vectors is mediated by transient plasmid transfection of mammalian cell lines (293T or HeLa), or recombinant baculovirus infection of insect cell lines (Sf9 or Ao38). Plasmids or recombinant BEVs encode the minimally required AAV and/or Adenovirus elements to promote rAAV vector assembly, genome replication, and packaging. Ultimately, rAAV vector particles are harvested and/or purified for downstream applications. The simplicity of rAAV vector genomes (encompassing only the AAV2 *ITRs* and promoter controlled transgene) and nature of minimally required AAV/Adenovirus factors also highlights the importance of host-AAV interactions and their impact on rAAV vector production *in vitro*. The inclusion of minimal viral proteins in rAAV vector production systems indicates that an array of cell-intrinsic factors need exploiting in order to propagate intact and infective rAAV vectors. A potential cell-intrinsic factor, Y-Box protein (YB)1, is thought to negatively regulate rAAV vector processing (Satkunanathan *et al.*, 2014). Human YB1 is a cold-shock domain (CSD)-containing protein, that harbours a significant array of functions: promoting cell survival (Lasham *et al.*, 2003; Homer *et al.*, 2005), cell proliferation (Swamynathan *et al.*, 2002; Fujiwara-Okada *et al.*, 2013; Wang *et al.*, 2015b), DNA repair (Gaudreault *et al.*, 2004; Das *et al.*, 2007; de Souza-Pinto *et al.*, 2009; Kim *et al.*, 2013), gene transcription (Stickeler *et al.*, 2001; Wei *et al.*, 2012), protein translation (Davydova *et al.*, 1997; Pisarev *et al.*, 2002; Nekrasov *et al.*, 2003), and responds to a number of stress responses (Ohga *et al.*, 1998; Shibahara *et al.*, 2004; Das *et al.*, 2007; Somasekharan *et al.*, 2015).

It was previously identified that human YB1 protein associates with rAAV vector particles and negatively regulates rAAV vector production in 293T cells (Satkunanathan *et al.*, 2014). Given this, the work presented in this thesis aimed to utilise the CRISPR/Cas9 genome editing system to regulate endogenous gene expression of human YB1 or Y-Box protein homologues (SfYB) in AAV vector producer cell lines – 293T and Sf9, respectively. We initially and extensively characterised knockout cell lines using a wide repertoire of assays to evaluate the genotype and phenotype of knockout cell lines. In doing so, we demonstrated a clear advantage of single cell cloning of CRISPR/Cas9-transfected 293T. This resulted in clones that were homozygous for the knockout mutation. Additionally, we emphasise the importance of using target-specific antibodies or CRISPR-validated antibodies in Western blotting for knockout cell lines using CRISPR/Cas9 genome editing.

However, YB1 knockout sensitised cell lines to chloroquine-induced cell death, and in turn restricted the use of this cell line for rAAV vector production by triple transfection with Calcium Phosphate precipitation method supplemented with chloroquine. Upon further analysis, rescue and/or resistance to this chloroquine-induced cytotoxic phenotype were achieved by the restored expression of YB1 and mutated YB1 encompassing mainly the CSD. A complex interplay between YB1's alanine-proline rich (A/P) domain or C-terminus domain (CTD) with the CSD seemed to regulate the rescue/resistance from chloroquine-induced cytotoxicity. Here, the general consensus was drawn that the CSD was the key mediator of the resistance phenotype to chloroquine-induced cytotoxicity. It was noted that the CTD alone (encompassed by YB1 Δ 5) conferred a rescued phenotype, however; therefore, whether or not the CTD, CSD, and/or A/P domain regulate each other in this function remains to be characterised. This is difficult to predict given the intrinsically disordered nature of YB1's termini as explored in Chapter 5. However, this is not unreasonable to assume given that YB1 forms oligomers of itself, especially via its CSD, of which process is modulated by the termini sequences (Guryanov *et al.*, 2012). Furthermore, mutation analyses showed that DNA binding activity that is exhibited by the CSD is influenced by the CTD (Chen *et al.*, 1995).

Alternative methods of rAAV vector production were unaffected by the YB1/SfYB knockout phenotype, and included triple transfection using PEI- or infection with recombinant baculovirus- based methodologies. As a result, comparable levels of AAV Rep and Vp1-3 expression and vector genome titres were calculated between control and knockout cell lines. Lastly, direct associations between AAV and human YB1 were inferred using DNA-YB1 affinity

pulldown assays and confocal laser scanning microscopy. Potential interactions between human YB1 and AAV2 *ITR* sequence was successfully demonstrated by means of a putatively identified YB1 binding motif (5'-GGGGTT-3').

7.2 CRISPR/Cas9 genome editing efficiently generated stable knockout cell lines

We demonstrated that YB1 and SfYB knockout cells were established using CRISPR/Cas9 genome editing technology, targeting *YBX1* gene at exons 1, 5 or 7, and intron 6, and *SFYB* at exon 1. The *YBX1* gene is located on chromosome 1 (Ch1p34.2), and it so happens 293T cells exhibit a pseudotriploid karyotype (Bylund *et al.*, 2004; Kim *et al.*, 2009), and 293T cells likely harbour three sets of the *YBX1* alleles. Therefore, the efficiency of targeting *YBX1* is complicated given multiple copies of the *YBX1* gene. Despite the pseudotriploid nature of 293T cells, potential YB1 knockout cell lines, particularly single cell clones, were successfully expanded. Knockout efficiencies were 20% for *YBX1*sgRNA1 and 60% for *YBX1*sgRNA3 designs. However, when taking into consideration all clones analysed, overall 20% knockout efficiency was achieved. This was attributed by the dominant phenotype of a number of clones – partial YB1 knockout. However, we found that despite the difficulty expected in targeting multiple copies of the *YBX1* gene, that CRISPR/Cas9 genome editing benefitted from single-cell cloning. This benefit was also emphasised by CRISPR/Cas9 genome editing of Sf9 cells, which could not be established as a monoclonal cell population. A consequence of this practical limitation was that the knockout phenotype was eventually lost due to the growth defect observed by SfYB-disrupted Sf9. Also, Sf9 cells exhibit a polyploidal karyotype, with diploid cells demonstrating chromosomal instability and propagating as tetraploidy over time in culture (Jarman-Smith *et al.*, 2002). The expansion of wildtype *SFYB* in this manner and the bottleneck generated by slower growing SfYB-disrupted cells likely eventuated in the recovery of the parental Sf9 population.

However, it is appreciated that CRISPR/Cas9-mediated knockout efficiency was reported to be insensitive to the copy number of the target gene, but that was determined instead by the efficiency at which sgRNAs recognise and bind to its target sequence (Yuen *et al.*, 2017). This was of important consideration when targeting sequences within cancer cell lines that exhibit considerable cytogenetic abnormalities and genomic instabilities that result in aneuploidy (Solomon *et al.*, 2011; Zhao *et al.*, 2014; Frattini *et al.*, 2015). In the present study, the transient nature of Cas9 expression, and in turn sgRNA from the pCRISPR-*YBX1*sgRNA(1-3) constructs and sgRNAsf1:rCas9 RNP transfections infer a transient period in which

CRISPR/Cas9 mediates DSBs at the targeted sequences. This was in contrast to Yuen *et al.* (2017), whom produced stable Cas9-sgRNA expressing cell lines to determine knockout efficiencies of the integrated *EGFP* transgene. Therefore, the persistent nature of Cas9-sgRNA expression off-sets the limitation exhibited by transient expression, and ultimately enhances the occurrence of targeted knockouts.

YB1 knockout clones A2 and C5, and to a lesser extent clone B2, exhibited stable YB1 knockout phenotype even up to 12 months in culture. This was particularly reflective of sufficiently established single cell clones, but also demonstrated a clear benefit of the genome editing tool over alternative gene manipulation tools. For example, siRNA can demonstrate targeted knockdown for up to 10 days post-transfection (Bartlett and Davis, 2006). Whereas shRNA technology, which, despite being able to stably introduce these elements into the target genome with lentiviral vectors or generate transgenic mice, can exhibit transient knockdown phenotypes (Satkunanathan *et al.*, 2014; Ajiro *et al.*, 2015). Transduction of shRNA-encoding lentiviral vectors function to incorporate the DNA intermediate of their vector genome into the target genome. Although, the process of integration site selection is rather non-specific, but does show preferential integration into active gene loci (Gierman *et al.*, 2007). The integration location partly defines the transcriptional activity of the transgene(s) encoded by the lentiviral vector, and/or the endogenous genes in which the integration has occurred (Hargrove *et al.*, 2008; Moiani *et al.*, 2012). On the other hand, CRISPR/Cas9 genome editing is precise and specific, providing the potential for off-target sequences is minimised. There is also the consideration that CRISPR/Cas9 genome editing incorporates only indels, or desired mutation with a donor template, at the specific target site. Inasmuch, minimum changes to the genomic blueprint and DNA composition is made using CRISPR/Cas9 genome editing, compared to alternative methods such as shRNA technology.

7.3 Complex spectrum of mutation profiles, alternative splicing motifs, and splicing activation are consequences of CRISPR/Cas9 genome editing

Analysis of sequencing profiles of the CRISPR/Cas9-targeted regions of YB1 knockout clones and SfYB-disrupted cells revealed a spectrum of mutations, mainly deletions, as a result of CRISPR/Cas9 genome editing. SpCas9 endonuclease generates DSBs at the target sequence, precisely 3nt upstream of the adjacent PAM 5'NGG3' (Gasiunas *et al.*, 2012). After which, endogenous NHEJ DNA repair pathway recognises the DSB for error-prone DNA repair and mutagenesis (Bibikova *et al.*, 2002). However, NHEJ is typically accompanied by DNA resection

at the site of DSB, and can include the loss >25bp (Su *et al.*, 2016b; Shin *et al.*, 2017; Sorenson *et al.*, 2017) prior to DNA polymerase-mediated repair. Large resection of DNA at DSBs and subsequent error-prone DNA polymerase activity may explain the localisation of the 1bp deletion that occurred 104bp upstream of the targeted PAM for the YB1 knockout clone A2. This was in contrast to the mutations observed for B2 and C5 clones, which reside well within the confines of their respective targeted sequences, and prevents the same locus from being re-targeted by CRISPR/Cas9 and *YBX1*sgRNA-2 or -3, respectively.

Furthermore, given that the selected YB1 knockout clones were derived from single cell clones and 293T exhibit a pseudotriploid karyotype (Bylund *et al.*, 2004; Kim *et al.*, 2009), the genotypes of the disrupted *YBX1* genes were of particular interest. We show evidence of clean chromatograms and Surveyor® mutation profiles indicative of homozygous genotypes for clones A2 and B2, for their targeted *YBX1* gene sequences. This was despite the error-prone nature of NHEJ. Additionally, this was in complete contrast to the genotype and mutation observed for clone C5 – an A>C substitution followed by an 8bp deletion, but associated with a complicated chromatogram indicative of a heterozygous genotype, with mutations in each allele. This is further supported by Surveyor® mutation screening and HRM curve analysis. Where unilateral targeting of multiple gene copies to produce the same CRISPR/Cas9-mediated mutation can be difficult, recent reports have described the ability to target multiple (up to 62 copies) of the porcine endogenous retrovirus in pigs using CRISPR/Cas9 (Yang *et al.*, 2015; Niu *et al.*, 2017). Of important note was the ability to distinguish that the heterozygous genotype was found to be in context of mutated alleles only, and in the absence of wildtype alleles by HRM curve analysis. The importance of distinguishing mutated alleles from wildtype alleles was necessary as the presence of wildtype alleles could promote the reversion of knockout between the mutation-containing chromatid(s) and wildtype sister chromatid(s), if present (Zaboikin *et al.*, 2017). Of course, similar was observed in genome edited Sf9 cells, where monoclonal populations could not be established, and a heterozygous genotype with intact wildtype *SFYB* gene eventually recovered.

Furthermore, the mutations that occurred across *YBX1*'s intron 4:exon 5 splice junction (B2 clone) potentiated in a disrupted wildtype splicing acceptor motif. Using the HSF v3.0 tool (Desmet *et al.*, 2009), predictive assessments for alterations in splicing motifs and alternative splicing were calculated, and compared against wildtype scorings as reference. This tool was more confidently harnessed when exact mutations were known. C5 clone's heterozygous

genotype and analysis of its Sanger sequencing results did not provide sufficient detail to accurately model the mutations for potential splicing alternations. However, clone B2 demonstrated a homozygous mutation at its intron 4:exon 5 splice junction, despite exhibiting a partial YB1 knockout phenotype. In this particular case HSF v3.0 predicted probable changes in splicing motifs, with alternative splicing (prediction 2) potentiating in a reading frame that was still in-frame with the wildtype stop codon, but encoded a truncated YB1 protein that shared 99% sequence homology. The truncation was a result of a 3aa deletion corresponding to residues 119-121 (Gly-Ala-Glu) of wildtype YB1. These residues comprise a part of YB1's CSD (YB1 CSD is mapped to between residues 51-129). This would explain the partial YB1 knockout phenotype observed as alternative splicing of an in-frame, albeit truncated, YB1 would be still be recognised by N- and C- terminus-specific α YB1 used in the study. Although it remains to be elucidated whether or not the truncated YB1 protein retains all, if any, wildtype function despite the 3aa deletion. The 3aa deletion does not coincide with YB1's CSD RNP binding domains (RNPI and RNPII), which follow the consensus sequences: Asn-Gly-Tyr-Gly-Phe-Ile and Val-Phe-Val-His-Phe, respectively (Landsman, 1992; Kljashtorny *et al.*, 2015). However, the defining β -barrel secondary structure of the CSD and its ability to bind to nucleic acids (Ladomery and Sommerville, 1994; Izumi *et al.*, 2001; Kljashtorny *et al.*, 2015) may be compromised. Alternatively, the truncated protein could function in a dominant-negative manner, and interfere with data analysis and the interpretation of results, essentially confounding downstream investigations (Kaphahnke *et al.*, 2016).

The significant homology shared between wildtype YB1 and the 3aa truncated variant led us to conclude that the B2 clone may not be a suitable YB1 knockout model. At least not without fully characterising the expressed protein, further. This clone also reflects the importance of understanding the exact mutation profiles introduced by CRISPR/Cas9 genome editing, and the potential impact mutations can have on alterations in splicing motifs and alternative splice site activation. Ultimately, A2 and C5 presented as the most promising and confident YB1 knockout clones generated by CRISPR/Cas9 genome editing of 293T cells.

7.4 Safety profile of YB1 knockout cell lines

The CRISPR/Cas9 genome editing tool derives mainly from *S. pyogenes* type II CRISPR-Cas system, and more importantly the associated SpCas9 endonuclease, therefore, has a prokaryotic origin. *S. pyogenes* is a common human commensal, typically contributing to the skin microbiota. The Gram-positive streptococci are known to be pathogenic and cause

pharyngitis, scarlet fever, and necrotising fasciitis (Medina *et al.*, 2003; Terao *et al.*, 2008). Much of the human population displays sero-positivity towards *S. pyogenes* infection, with as much as 100% of the adult population displaying protective antibodies and cell-mediated immune responses against it (Mortensen *et al.*, 2015). Therefore, the possibility of pre-existing immunoreactivity towards Cas9 protein presents a potential hurdle for its utilisation in cellular or gene therapy applications. One of the main purposes of establishing an YB1 knockout cell line was to engineer a novel cell line for rAAV vector production, principally due to the previous observation that YB1 knockdown by shRNA technology correlated with enhanced rAAV vector titres (Satkunanathan *et al.*, 2014). This warranted the deep and extensive characterisation of the YB1 knockout cell lines generated by CRISPR/Cas9 genome editing; but also included investigating the safety profile of using these cell lines by examining for stable Cas9 endonuclease expression.

Long-term follow-up (up to 12 months) of YB1 knockout clones in culture demonstrated that the pCRISPR-YBX1sgRNA1-3 transfections of 293T exhibited transient transfection dynamics. Cas9 was not detected in YB1 knockout clones that were cultured for up to 12 months, indicating that Cas9 and sgRNA expression occurred in a transient nature; at least by the limits of detection by Western blotting. Therefore, pCRISPR plasmid integration into the target genome was not a risk that was observed. This was in spite of the fact that plasmid integration into the host genome can occur *in vitro and in vivo* (Chen and Okayama, 1987; Wang *et al.*, 2004b; Schiedner *et al.*, 2008); albeit, at near negligible frequencies *in vitro* without an antibiotic selection pressure to promote isolation of transformed cell lines. In fact, the pCRISPR constructs possessed no selectable antibiotic marker, and transfection followed by single cell cloning was the study design of choice to clonally isolate YB1 knockout cell lines.

Ultimately, Cas9 stable expression was not detected in selected YB1 knockout clones. This characterisation potentially increased the safety profile of the engineered 293T cells, especially in context with rAAV vector production. We currently understand, and observed, that rAAV vectors package or associate with cell-intrinsic constituents. Packaged host cell proteins have been identified in association with rAAV vectors, and include a range of host cell factors: nucleolin, nucleophosmin, YB1, RuvB, CypA, annexin V (Dong *et al.*, 2014; Satkunanathan *et al.*, 2014). The exact relevance of these packaged host cell factors to the biology of AAV and AAV processing has not been fully elucidated. In a CRISPR/Cas9 engineered rAAV producer cell line, the production process could include the incorporation of Cas9

endonuclease into assembling rAAV virions. In turn, the mis-packaged Cas9 could be transferred along with the desired transgene into target cells/tissues, and influence immune responses. In fact, Adenovirus-CRISPR and AAV-CRISPR systems have been examined *in vivo* for immune responses in mice models (Wang *et al.*, 2015a; Chew *et al.*, 2016). In the latter instance, AAV-split-CRISPR involved rAAV9 vectors encoding Cas9 to target the *Mstn* gene for editing *in vivo*. However, delivery of AAV-split-CRISPR or DNA encoding full length Cas9 elicited inflammatory immune reactions, antigen-specific T-cell activation and expansion, and generation of Cas9-specific antibodies in a Cas9-dependent, AAV9-independent manner (Chew *et al.*, 2016). More recently and more relevantly, pre-existing immune responses to SpCas9 and SaCas9 was identified from human blood serum. This was in the form of Cas9-specific, immunoreactive T-cells (T_H1 and cytotoxic CD8⁺ T-cells) and production of cytokines IFN γ and TNF α in response to Cas9-antigen stimulation (Charlesworth *et al.*, 2018). Therefore, careful CRISPR/Cas9 designs may have to be considered for genome editing of human cells for gene therapy. However, in the present study, the level of Cas9 expression was undetectable by Western blotting in YB1 knockout clones. Cas9 protein is unlikely to be packaged and mediate Cas9-associated immunoreactivity, for downstream applications, particularly gene therapy applications. In turn, the YB1 knockout cell line potentiates a risk-free model for rAAV vector production.

7.5 Baculovirus technology for *S. frugiperda*-native recombinant protein expression

Insect cell-based systems have proven extremely useful in the production of large quantities of recombinant protein. The Sf9 and Ao38 (or High Five™) cell lines are key examples of systems that are used for the heterologous expression of recombinant proteins (Lopez-Vidal *et al.*, 2015; Ge *et al.*, 2016; Mohseni *et al.*, 2016). This also includes the scale-up of rAAV vectors (Carpentier *et al.*, 2012). Insect cells are generally infected with baculoviruses, given their high and robust efficiency, productive infectivity (Guo *et al.*, 2005), and strong baculovirus-derived promoters (Smith *et al.*, 1983b; Martinez-Solis *et al.*, 2016). Whereas, transfection of Sf9 cells (and insect cell lines in general) with naked plasmid DNA is typically associated with poor to moderate transfection capacities, often ranging between 30-50% (Maeda *et al.*, 2005; Ogay *et al.*, 2006). We observed a similar range of transfection efficiency using Cellfectin™ II reagent and a GFP reporter plasmid. However, regardless of the gene transfer method applied, a discrepancy in intracellular His-tag staining was observed. This was best exemplified by baculovirus infections using BacGFP(His)₁₀, in that despite high infection efficiency a lower percentage of the cell population stained positively for the His-tag. It was noted that the significant disorder in the N- and C- termini in Y-Box proteins (Guryanov *et al.*, 2012), including

SfYB may impact on the exposure of the His-tag of rSfYB(His)₁₀. Ultimately it was concluded that intracellular His-tag staining was not an accurate correlate for His-tag expression for the target recombinant protein, in the current studies.

We also demonstrated a comparable expression profile between infected Sf9 and Ao38 cell lines for the rSfYB(His)₁₀ protein expression, as determined by Western blotting. Although, it could be argued that some relatively enhanced expression of the rSfYB(His)₁₀ was observed in Sf9 cell line. This came as a surprise given multiple records that suggest *T. ni*-based cell lines (High Five™ and Ao38) were considerably better producers of recombinant proteins compared to Sf9 or Sf21 (Wickham and Nemerow, 1993; Wilde *et al.*, 2014). This is presumably due to the higher demands in metabolism and nutrition during baculovirus infections of *T. ni*-derived cells (Rhiel *et al.*, 1997; Pushparajan *et al.*, 2017). However, we take note that the target rSfYB(His)₁₀ was cloned from the endogenous SfYB protein that was derived from *S. frugiperda* (and Sf9 cell line). Therefore, it was recognised that the recombinant protein was expressed from its endogenous cell system. It could be reasoned that the expression of recombinant protein in their endogenous cell systems is advantageous. Although, it is difficult to define how so in the current study, but it can be ruled out that the improved production dynamics in Sf9 compared to Ao38 was not a result of a transcriptional advantage. This was because rSfYB(His)₁₀ expression was under the promoter control of baculovirus-specific p10 promoter, and not its native promoter. Therefore, translational regulation of the rSfYB(His)₁₀ was the likely process affected.

As per keeping of a eukaryotic cell line, Sf9 cells are capable of post-translational modifications of recombinant proteins expressed from baculoviruses (Vrljic *et al.*, 2011). These are usually necessary for full functionalities of recombinant proteins derived from heterologous gene expression. Post-translational modifications was considered a likely occurrence of rSfYB(His)₁₀ expression, given the electrophoretic mobility shift from the predicted 29.2kDa to approximately 50kDa by SDS-PAGE. The exact post-translational modifications were not defined in the present studies for SfYB, but select modifications and their cross-talk are capable of regulating protein degradation. For example, protein kinase C (PKC) α is phosphorylated to promote its sumoylation and protein stability, which in turn protects against ubiquitination and proteasomal degradation (Wang *et al.*, 2015d). Endogenous SUMO proteins are components of Sf9 lysates (Langereis *et al.*, 2007), but were not implicated in sumoylation of exogenously expressed products (mammalian in origin) from baculoviruses

tested (Langereis *et al.*, 2007). Therefore, the advantage of post-translational modifications is likely a feature we observe in the present study that permitted improved rSfYB(His)₁₀ expression in Sf9 cells compared to Ao38.

Nonetheless, antiserum raised against the rSfYB(His)₁₀ was generated and proved a highly reactive and specific for SfYB in Sf9 cell lines. This was warranted given the reduced homology shared between human YB1 and the newly identified SfYB (predominant homology was calculated in the CSD, and extensive variation was reported for the remainder of the whole molecule). Additionally, the current pool of α YB1 available commercially were mostly raised against human YB1 peptides, but were still capable of cross-reactivity for SfYB in Sf9 lysates. Even with C-terminus specific α YB1, where homology and conservation between YB1 and SfYB was lowest, was found to cross-react almost comparably to the N-terminus specific α YB1. The fact that polyclonal antibodies are raised against several different epitopes of an antigen target compared to single, defined epitope derived monoclonal antibodies, proves advantageous for cross-reactivity. However, antibodies raised against human specific antigens do not always cross-react with protein homologues in different species or isoforms (Werther *et al.*, 1996; Rosenbluth *et al.*, 2009). Therefore, with the generation of rSfYB(His)₁₀-specific antiserum, we conclusively identified the Y-Box protein (SfYB) in *S. frugiperda*, Sf9 cell line.

7.6 SfYB is a *S. frugiperda* Y-Box protein

Y-Box proteins encompass a large family of CSD-containing proteins that are evolutionarily conserved, and essentially found encoded by all domains of life (Karlson and Imai, 2003; Nakaminami *et al.*, 2006). The CSD is homologous to prokaryotic cold-shock proteins, which function in response to a cold shock. The response functions to regulate transcriptional and translational processes to maintain prokaryotic cell viability (Wouters *et al.*, 1999). We identified human Y-Box protein homologues that were highly conserved in *Spodoptera spp.* and *T. ni*. Analyses demonstrated conservation and homology defined to the sequences that correspond to the CSD. This comes as no surprise because evolutionary conservation of the cold-shock proteins and the homologous CSD has been consistently reported (Mani *et al.*, 2012; Kljashtorny *et al.*, 2015). In fact, *E. coli*'s CspA and human YB1's CSD share approximately 43% sequence homology (Wang *et al.*, 2000b). The increased homology observed between human YB1 and insect homologues (approximately 84%) was likely a matter of evolutionary divergence from unicellular cold-shock proteins, of which required a conserved functionality for eukaryotic adaptation and development. For instance, it has been reported that

mammalian YB1 is indispensable in late embryonic development of mice (Lu *et al.*, 2005; Uchiumi *et al.*, 2006). It remains unclear if the highly conserved CSD or evolutionarily acquired A/P or CTD domains, or a complex interplay, promotes progressive embryonic development. Of course, the expansion of the N- and C- termini of eukaryotic Y-Box proteins are acquired features, but showed the most variation between species of insects examined in the current study.

It was readily acknowledged that the characterisation of insect Y-Box protein homologues of the *Lepidoptera* order is rather limited (Takiya *et al.*, 2004; Nishita and Takiya, 2005). Human YB1 is predicted 36kDa in molecular weight but resolves to approximately 50kDa by SDS-PAGE. The SfYB protein revealed a similar electrophoretic mobility (approximately 50kDa) despite being predicted 29.2kDa. However, *B. mori*'s BYB was first reported to resolve to approximately 36kDa (Takiya *et al.*, 2004) or approximately 50kDa (Nishita and Takiya, 2005). The extent of BYB's post-translational modifications (if any) has not been elucidated, and may likely contribute to shifts in the electrophoretic mobility of the protein (Shirai *et al.*, 2008). This was of interest because human YB1's function has been reported to be regulated by post-translational modifications. These include: phosphorylation of S102 – implicated in breast cancer oncogenesis and chemoresistance (Stratford *et al.*, 2008; Dhillon *et al.*, 2010), S165 or S176 – regulating NFκB activation and NFκB-induced genes, and implicated in colon cancer (Prabhu *et al.*, 2015; Martin *et al.*, 2017), or ubiquitylation for targeted 20S proteasome-mediated degradation (Lutz *et al.*, 2006), to name a few. Therefore, given a shift in electrophoretic mobility, it is likely SfYB is as extensively regulated in its function(s) similarly to human YB1. The context of which these modifications occur does require further elucidation, as the aforementioned post-translational modifications of YB1 were mainly characterised in context to carcinogenesis. Treatment of rSfYB protein with phosphatases or immunoprecipitation with poly-ubiquitin antibody may reveal the exact post-translational modifications for further scrutiny.

SfYB also demonstrated structural similarities to human YB1 and to the BYB controls. It was deduced that generally the homologues were composed of three main functional domains – a short N-terminal domain, the conserved CSD, followed by a CTD. This was similar to BYB (Takiya *et al.*, 2004) and *Philosamia ricini* Y-Box protein (PYB) homologue (Mani *et al.*, 2012). Furthermore, secondary structures were recognised by predictive analyses in the conserved CSD, including five β-strands assembling into a β-barrel. This was synonymous to human YB1

(Kloks *et al.*, 2002; Kljashtorny *et al.*, 2015) and *E. coli*'s CspA (Newkirk *et al.*, 1994). Interestingly, near complete matches in 3D-structure and folding were predicted between superimposed BYB, SfYB or TnYB CSDs to human YB1 template model 5yts, as per TM-scores of approximately 0.94-97. It is appreciated that tertiary structures of published Y-Box protein homologues using, for example – comparative 3D-modelling of YB1 and *Bacillus caldolyticus* CspB (a bacterial cold-shock protein) generated an RMSD value of 0.49Å using molecular dynamics simulations (Kljashtorny *et al.*, 2015). However, comparative analysis by RMSD is considered a poor measure of protein structure conservation. The RMSD score is largely influenced by the overall length of the alignments. Therefore, large protein sequence alignments generate a large RMSD value, of which is not scaled for a relative measure of structural conservation (Maiorov and Crippen, 1994). The TM-score overcomes this dependence on the length of alignment by weighing the distances between superimposed residues more strongly compared to those further apart (Zhang and Skolnick, 2004). Therefore, the proposed high TM-scores between the 5yts model and Y-Box protein homologues indicated that the modelled proteins reflect accurate predictions.

Nonetheless, the CSD is involved in DNA- and RNA- binding activities (Ladomery and Sommerville, 1994; Izumi *et al.*, 2001; Kljashtorny *et al.*, 2015), and the RNP-I and RNP-II motifs were successfully identified as generally well conserved (Landsman, 1992) in the CSDs of BYB (Takiya *et al.*, 2004), SeYB, SIYB, SlittoYB, SfYB and TnYB homologues. In fact, BYB and the structurally conserved PYB showed similar preferential affinity or binding to ssDNA and ssRNA, and less so to dsDNA species (Nishita and Takiya, 2005; Mani *et al.*, 2012), or enhanced DNA binding of mammalian transcription factors for enhanced transcriptional activities (Takiya *et al.*, 2004). Despite these functions being mediated mostly by the CSD, the flanking domains of BYB or PYB were considered supportive of specific ssDNA binding (Takiya *et al.*, 2004; Mani *et al.*, 2012). This also seems to be the case for human YB1 (Izumi *et al.*, 2001; Tanabe *et al.*, 2015), and is perhaps likely for the newly characterised SfYB. However, experimental evidence is required to completely classify these homologues as orthologues.

The disruption of SfYB was also associated with a significant growth defect when compared to parental Sf9 cells. Human YB1 is well characterised to have direct associations with cell cycle progression and cell proliferation (Fujiwara-Okada *et al.*, 2013; Kotake *et al.*, 2013; Wang *et al.*, 2015b; Dey *et al.*, 2016). The knockdown of YB1 protein expression was associated with cell cycle defects and slowed cell proliferation, even in selected tumour cell lines (Shiota *et al.*,

2008; Wang *et al.*, 2015b). YB1 directly associates and downregulates the tumour suppressor gene, *p16^{INK4A}* locus in order to promote replication (Kotake *et al.*, 2013). The encoded cyclin kinase inhibitor regulates cyclin D/cdk4 kinase activity, which would otherwise lead to the phosphorylated pRb and accumulation of E2F for entry into S-phase of the cell cycle (Kato *et al.*, 1993; Ezhevsky *et al.*, 1997; Narasimha *et al.*, 2014). However, a growth defect was not observed in YB1 knockout cell lines established by CRISPR/Cas9 genome editing. The immortalisation of 293T cell lines by Adenovirus type 5 *E1* ORF (Graham *et al.*, 1977) and SV40 large T antigen (DuBridge *et al.*, 1987) likely circumvents this growth inhibition. The products of the *E1* ORF, particularly E1A, promotes host cell entry into active S-phase of the replicative cell cycle by up-regulating cyclin E and cdc25A expression and activity (Spitkovsky *et al.*, 1996). E1A is also directly implicated in the pRb/E2F-1 pathway, by binding to the pRb family of proteins to release key transcription factors that are defined as key regulators of S-phase entry from the G1 stage of the cell cycle (Nevins, 1990).

Nonetheless, given the homology shared between SfYB and human YB1, it is likely that certain functions may also be shared – including control and regulation of the cell cycle progression and proliferation. The inability to establish single-cell clones of the heterozygous SfYB-disrupted cell line and growth defect, however, may have resulted in a bottleneck for population density between SfYB-disrupted cells and parental Sf9 after the establishment of the Sf9ex1.7/10X cell line. Meaning the parental Sf9 cells would eventually outgrow the slower growing SfYB-disrupted cells. This was eventually observed in the current study, with later passaged Sf9ex1.7/10X cell line exhibiting SfYB protein expression restored to wildtype levels. Ultimately, the target gene of interest was amenable to genome editing by CRISPR/Cas9 technology in Sf9 cells; however, the likely role of the SfYB protein in cell cycle progression and proliferation meant the gene edited cell line was eventually lost by outcompeted growth.

7.7 Unprecedented consequences to CRISPR/Cas9 genome editing

Significant and enhanced cell cytotoxicity phenotype was observed for YB1 knockout clones treated with chloroquine in any instance. This cell cytotoxicity phenotype was generally considered independent of the Calcium Phosphate mix, transfection of plasmid(s), and transfection efficiencies for the input pAAV2-hrGFP reporter plasmid. Despite the reported cytotoxicity exerted by AAV2 Rep78 protein expression (Yang *et al.*, 1994; Schmidt *et al.*, 2000) YB1 knockout cell lines were concluded to be especially sensitive to the effects of chloroquine. Lastly, PEI-based transfection or baculovirus infections for rAAV vector production showed

comparable cytotoxicity profiles between 293T and YB1 knockout cells, and Sf9 and SfYB-disrupted cell lines, respectively.

It was noted that rAAV production by Calcium Phosphate precipitation method has shown to involve the release of vector into the culture media (Vandenberghe *et al.*, 2010), and PEI-mediated production processes have also reported rAAV vectors appreciably within the culture media (Lock *et al.*, 2010). Some reports describe 293T cells rounding up and detaching from the plate after transfection with Calcium Phosphate precipitation or PEI (Lock *et al.*, 2010; Huijun *et al.*, 2014). With these in mind and given AAV vectors are non-enveloped, some level of cytotoxicity may be a direct consequence of rAAV production, resulting in the release of particles into the culture media. However, the significant effect from chloroquine treatment alone on YB1 knockout cells suggested that chloroquine may be functioning in a more adverse and dominant manner to this particular CRISPR/Cas9 genome engineered cell line or clone. Perhaps the loss of YB1 has dire consequences that would otherwise be protective to chloroquine treatment?

Chloroquine is a quinolone compound and weak base, which has been extensively reported to supplement the Calcium Phosphate precipitation method for transfection (Luthman and Magnusson, 1983; Hasan *et al.*, 1991; Felgner *et al.*, 1994; Kariko *et al.*, 1998). Chloroquine aids transfection in two main ways, i) chloroquine neutralises acidic pH in endocytic vesicles to prevent further degradation of input exogenous DNA, and ii) chloroquine also has been shown to displace polycations from input DNA to promote nuclear shuttling (Cheng *et al.*, 2006). Alternatively, chloroquine is also well recognised as a first-line treatment against malaria due to infection with *Plasmodium* strains (Watt *et al.*, 1988; Levy *et al.*, 1991; Pillai *et al.*, 2001). Chloroquine is toxic to *Plasmodium* strains, and is thought to function by accumulating in the digestive vacuoles of *P. falciparum* infected erythrocytes and interferes with the parasites ability to detoxify its digestive vacuole from the haemoglobin proteolysis by-product, ferriprotoporphyrin IX (Bray *et al.*, 1999). This leads to significant damage and disruption to the *Plasmodium* membranes (Ginsburg and Demel, 1983; Zhang and Hempelmann, 1987). It has also been reported protective against Zika virus infection *in vitro* (Delvecchio *et al.*, 2016), and *in vivo* using SJL mice models that demonstrated reduced vertical transmission incidences (Shiryaev *et al.*, 2017). However, chloroquine-mediated toxicity and cell death of uninfected Vero cells was detected, albeit, at high working concentrations of chloroquine (100µM and above) than used in the present study. Although a working chloroquine concentration of 25µM

was used in the present studies, 293T and CRISPR/Cas9-genome edited derivatives were principally examined given their relevance in rAAV vector production, and not Vero cells. Therefore, the use of different cell lines may account for the relatively enhanced resistance of Vero cells to chloroquine cytotoxicity, compared to 293T and YB1 knocked out-293T derivatives.

On the other hand, PEI is a linear, cationic polymer of aziridine monomers, and functions to bind to the phosphodiester backbone of DNA or directly to DNA grooves, and condense it into positively charged particles that are capable of binding to cell surfaces (Utsuno and Uludag, 2010). This permits DNA:PEI complexes to become receptive to internalisation by endocytosis, and input DNA is further protected by the buffering potential by PEI in acidic endosomes (Boussif *et al.*, 1995; Sonawane *et al.*, 2003). This involves binding and retaining of protons which influences the influx of Cl⁻, which perturbs the endolysosomal osmolality to release DNA into the cytoplasm for nuclear entry (Pollard *et al.*, 1998; Akinc *et al.*, 2005). Therefore, similarly to chloroquine, PEI buffers the endolysosomal pH to protect DNA degradation and enhance transfection efficiency. Given PEI operates to buffer the acidic pH of the endolysosomal compartment analogous to chloroquine, the present studies could indicate that an alternative pathway involving chloroquine that is associated with YB1, but independent of PEI, is responsible for an enhanced cytotoxic phenotype in YB1 knockout cells.

The cytotoxic potential of chloroquine has been well defined in a number of tumour cell lines (Zaidi *et al.*, 2001; Geng *et al.*, 2010; Kim *et al.*, 2010; Lakhter *et al.*, 2013; Frieboes *et al.*, 2014; Vessoni *et al.*, 2016). It has also been identified as a significant sensitising agent of cancer cell lines for targeted cell death (Verschooten *et al.*, 2012; Schmukler *et al.*, 2014; Park *et al.*, 2016). For example, Lakhter *et al.* (2013) demonstrated that chloroquine prevented the degradation of pro-apoptotic p53-upregulated modulator of apoptosis (PUMA) in melanoma cell lines in a lysosome-independent manner. This resulted in decreased cell viability and increased levels of caspase activity. Also, Kim *et al.* (2010) demonstrated that chloroquine stabilised p53, resulting in transcriptional upregulation of p53-regulated gene targets, including apoptotic *BAX*; although, this was largely dependent on wildtype p53 status. Therefore, there is a clear involvement of chloroquine in stimulating the apoptotic cell death pathway(s). In contrast to this, YB1 protein has been characterised as a dominant negative regulator *TP53* gene expression, and of p53 function and p53-mediated apoptosis (Lasham *et al.*, 2003; Zhang *et al.*, 2003; Homer *et al.*, 2005; Schitteck *et al.*, 2007). The overexpression of

YB1 in metastatic melanoma cell lines significantly correlated with chemoresistance, also (Schitteck *et al.*, 2007). Therefore, taking into account that chloroquine is capable of inducing apoptosis, along with YB1's role in negatively regulating p53-mediated apoptosis, a possible explanation to the heightened sensitivity of YB1 knockout cell line has been proposed. The sensitivity observed in response to treatment with chloroquine may be due, in part, to the lack of YB1 negatively regulating p53-mediated apoptosis induced by chloroquine. In parental 293T cells with intact YB1 expression and function, the adverse effect of chloroquine is most probably limited by a functioning relationship between p53 and YB1.

It is, however, noted that 293T cells are transformed 293 cells for the additional, constitutive expression of SV40 large T antigen (DuBridge *et al.*, 1987). The purpose of which was for the improved performance of replication and/or expression of transfected plasmids harbouring the SV40 origin of replication (DuBridge *et al.*, 1987; Heinzel *et al.*, 1988; Mahon, 2011). However, more importantly and with relevance to the YB1 knockout cell line sensitivity towards chloroquine, SV40 large T antigen has linked to disrupting p53's tumour-suppressing functions, and associated with promoting tumorigenesis (Zhen *et al.*, 1999). The large T antigen acts as a potent inhibitor of p53 tumour-suppressor functions. It directly binds to p53 tumour suppressor (Lilyestrom *et al.*, 2006), and in doing so, encourages the stabilisation of p53 levels that are thought to be functionally inactive or dysfunctional (Mietz *et al.*, 1992; Bocchetta *et al.*, 2008). This complex influences the biological function of p53 by promoting malignant cell growth and extended cell longevity (Lin and Simmons, 1991), possibly by activating the insulin-like growth factor-I signalling pathway – at least in primary human mesothelial cell (Bocchetta *et al.*, 2008). Additionally, the strong interaction between p53 and large T antigen prevents p53's ability to bind DNA sequences in a site-specific manner (Bargonetti *et al.*, 1992), influencing, in turn the cell's gene expression profile (Chang *et al.*, 1997; Hermannstadter *et al.*, 2009) in favour of cell transformation.

It is difficult to ascertain the interplay, if any, between SV40 large T antigen and YB1's anti-apoptotic functions or even chloroquine from the present studies and the current literature. It has been demonstrated that the human Polyomavirus JC virus large T antigen is able to interact with YB1 host cellular protein to positively modulate JC virus promoter activities (Chen *et al.*, 1995; Safak *et al.*, 1999). It is appreciated that JC virus shows a total 69% homology to SV40 (Frisque *et al.*, 1984). Thus, significant structural and functional conservation has been demonstrated between their large T antigens (Pipas, 1992; Wang *et al.*, 2004a), including its

capacity to bind to p53 (Staib *et al.*, 1996). The functional relationship between JC virus large T antigen and YB1 could extend to SV40 large T antigen, and may influence SV40 large T antigen-p53 tumour suppressor dysfunctional activities. By extension, this could affect the survival of 293T cells in the presence of chloroquine. This property may be further abrogated and lost in the established YB1 knockout cell lines, especially in the presence of certain cytotoxic stresses, including chloroquine.

Therefore, CRISPR/Cas9 targeting of host genes potentially introduces a limitation on inferences established from data analysis. This is especially of consideration when non-endogenous, viral factors used to establish immortalised cell lines (such as SV40's large T antigen) potentially show some interplay with the targeted gene and/or encoded protein(s). Altogether, this may warrant further investigations to elucidate any interplay between YB1 and SV40 large T antigen. Or at least encourage alternative transformation or immortalisation methodologies that minimally affect the blueprint of the host cell's genome, as is otherwise evident by integrating viral elements or plasmids. Perhaps, CRISPR/Cas9 genome editing could be implemented to generate a novel range of immortalised cell lines from primary cell lines. This could potentially off-set the limitations in data analysis and interpretations made on physiological protein function(s) that may in fact be an artefact due to the exogenous factors expressed to establish these immortalised cell lines.

The cytotoxicity associated with the YB1 knockout cell lines presented, may indicate that a baseline amount of YB1 is required to maintain select cell function(s) that in turn help with AAV production. This was perhaps observed with the knockdown of YB1 in 293T producer cells by Satkunanathan *et al.* (2014). However, YB1 knockout of MEF cell lines was sufficient to significantly enhance Dengue virus (DV) viral replication, production, and viral protein expression (Paranjape and Harris, 2007) – an alternative RNA virus. It is also noted that YB1 plays an important role in early development in mice, and YB1 knockout transgenic mice developed relatively normally up to embryonic day 13.5 (Lu *et al.*, 2005). After which, the YB1 knockout phenotype was associated with growth retardation and embryonic lethality (Lu *et al.*, 2005; Uchiumi *et al.*, 2006). Whereas, heterozygous mice (YB1^{+/-}) were phenotypically indistinguishable from the YB1^{+/+} counterparts, indicating a non-redundant role of YB1, at least in late embryonic development of mice. Given this observed embryonic lethality, the enhanced DV viral replication, production and viral protein expression observed by Paranjape and Harris (2007), may have been a product of the fact that YB1 knockout MEF lines were

generated and immortalised at embryonic day 13.5 (Lu *et al.*, 2005). A point at which the redundancy of YB1 on cell functions, which permitted embryonic development and additional unknown biological processes, remained intact. The loss of mouse YB1 in these YB1^{-/-} MEF lines may have contributed to a conducive cell line that enhanced DV vector processing. It is hard to ascertain whether or not YB1 knockout in MEF lines generated post-late stage embryonic development would recapitulate the enhanced DV replication and viral protein expression. In the current study, YB1 knockout cells were established using CRISPR/Cas9 genome editing of 293T cells. 293T cells were derived from immortalised 293 cells (Graham *et al.*, 1977), of which origin has not been fully established except derived from embryonic kidney tissue and thus foetal in origin (please refer to ATCC® CRL-1573™ [https://www.lgcstandards-atcc.org/products/all/CRL-1573.aspx?geo_country=gb]). It is not fully established whether or not 293 cells were established from tissue late in embryonic development – a point in which human YB1 could function non-redundantly, similarly to mouse YB1 in MEF lines by Paranjape and Harris (2007). Although, human embryonic kidney tissue is a far cry from MEF, it may be argued that the YB1 knockout genotype established in 293T cells was further uncondusive for rAAV vector production, given its non-redundant function in as of yet unidentified biological processes. This process(es) was perhaps otherwise maintained with a baseline level of YB1 expression in YB1 knockdown cell lines established by Satkunanathan *et al.* (2014) that permitted enhanced rAAV processing, albeit transiently.

Given this cytotoxicity association after CRISPR/Cas9-mediated knockout of human YB1, we were concerned that a similar pronounced effect could result after viral vector infections. Transductions with rAAV or lentiviral vectors did not significantly alter cell viability at all tested MOIs (MOIs: 10-100 for rAAV2GFP, and MOIs: 1-30 for lentiGFP vectors). Additionally, infections of Sf9 and SfYB-disrupted cell lines with baculoviruses (MOI: 3 each baculovirus for rAAV vector production) also demonstrated comparable cytotoxicity profiles. Therefore, the pronounced sensitivity to select transfection modalities thus far remained an isolated consequence of *YBX1* gene disruption by CRISPR/Cas9. Viral vector toxicity in YB1 knockout cells was a consideration due to the inherent, baseline toxicities that have been previously associated with viral vector transductions. For example, several rAAV vectors (serotypes 1, 2, 5-9) have been examined on primary cortical cultures derived from embryonic rat brain tissue and all rAAV serotypes, excluding AAV9, exhibited MOI-dependent cell toxicity (Howard *et al.*, 2008). Additionally, cell toxicities were found only partly attributed by the expression of the GFP reporter. The generation of rAAV serotypes encoding a mutated GFP persisted to exhibit

significant cell toxicity after cell infections, demonstrating that rAAV vector was responsible for the difference in cytotoxicity. Additionally, gene transfer using lentiviral vectors has been previously thought to elicit cell cytotoxicity, mainly due to the envelope protein component – VSV-G, which has been implicated in cytopathogenicity and increased caspase-3/7 activity as a result of VSV (serotype Indiana) infection of Vero cells (Hoffmann *et al.*, 2010).

The cytotoxicity profiles calculated after treatment with viral vectors was largely considered to be independent of transduction or transduction efficiency. The transduction efficiencies using lentiviral or baculoviral vectors were generally considered comparable between the corresponding cell lines using flow cytometry. Transductions of YB1 knockout cells with rAAV vector at MOIs 100 and 1000 were considered incomparable, however – almost 50% less than that in parental 293T. It has been previously shown that AAV are compartmentalised into Rab7A⁺ late endosomes and Rab11A⁺ recycling endosomes (Douar *et al.*, 2001; Ding *et al.*, 2006; Harbison *et al.*, 2009). AAV, in part, exploits the host cells' microtubule network to traffic itself towards the nucleus (Xiao and Samulski, 2012). YB1 has also been shown to interact with microtubules and stimulate their assembly *in vitro* (Chernov *et al.*, 2008). Additionally, YB1 was found to be required for centrosome maturation in metaphase cells (Kawaguchi *et al.*, 2015a), where the centrosome is essential for the nucleation and anchoring of microtubules. The polymerisation of microtubules occurs by a two-step process involving the nucleation of tubulin and then their elongation (Caudron *et al.*, 2002), which was largely promoted in the presence of YB1 (Chernov *et al.*, 2008; Kawaguchi *et al.*, 2015a). The disruption of endogenous YB1 can negatively affect microtubule assembly, as demonstrated by YB1 knockdown HeLa cells using siRNA reduced nucleation of microtubules from centrosomes (Kawaguchi *et al.*, 2015a), or complete loss of Rab11A⁺ recycling endosome localisation to centrosomes (Kawaguchi *et al.*, 2015b) were observed. In the case of the latter, Rab11A⁺ recycling endosomes are exploited by internalised AAV to traffic to the nucleus for efficient transduction or infection. However, it could be reasoned that YB1 knockout compromised rAAV transduction due to a potentially inefficiently established microtubule network and/or trafficking of Rab11A⁺ recycling endosomes. We observe only a reduction in transduction efficiency in YB1 knockout, and not the complete loss of transduction. This may imply that either AAV utilises additional means to traffic infected cells, or YB1 shows some redundancy for its role in microtubule assembly and centrosome maturation.

7.8 YB1 confers a protective phenotype against chloroquine reagent

YB1 is predominantly found in the cell's cytoplasm, with strict conditions that dictate its nuclear translocation, such as cell-cycle-dependent cues (Jurcrott *et al.*, 2003), genotoxic stresses (Zhang *et al.*, 2003), and UV irradiation (Koike *et al.*, 1997). The regulation of YB1's translocation, is mediated in part, by the presence of three nuclear localisation signals (NLS) cytoplasmic retention signals (CRS), and the 20S proteasome cleavage site that resides between residues 219 and 220, resulting in a truncated N-terminal YB1 fragment and a CTD fragment (Kim *et al.*, 2013). 20S proteasome-mediated cleavage of YB1 was generally a response to genotoxic stresses (Sorokin *et al.*, 2005; Kim *et al.*, 2013).

It seems that the interplay between NLS1-3 (found between 149-156aa, 185-194aa, and 276-292aa, respectively) and CRS1 and 2 between 52-101aa and 267-293aa, respectively (van Roeyen *et al.*, 2013), dictates the subcellular localisations of YB1 mutants. A list of subcellular localisations of full length YB1 and truncation mutants identified in the present study, and their characteristics pertaining to NLS, CRS, and the 20S proteasome cleavage site are listed in Table 4.1. Therefore, we observed that the presence of either CRS motifs dictated a cytoplasmic localisation of YB1 and mutants. These were in general agreement with previous reports (Bader and Vogt, 2005; van Roeyen *et al.*, 2013). Although, it was noted that van Roeyen *et al.* (2013) used an outdated annotation of the YB1 sequence (GenBank: J03827.1), which sports a shorter YB1 protein sequence (317aa) instead of 324aa (NP_004550.2). YB1 Δ 2 and YB1 Δ 4 were slight exceptions to this rule and instead showed some nuclear shuttling and localisation despite the presence of the CRS1 motif. Both YB1 Δ 2 and YB1 Δ 4 most closely recapitulates the genotoxic stress-induced 20S proteasome cleavage N-terminus fragment product, and therefore, it came to no surprise that their localisations was shared with the 20S proteasome cleavage product (Sorokin *et al.*, 2005; Kim *et al.*, 2013).

The chloroquine-induced cell death was reverted in YB1 knockout cells stably expressing YB1FL and select YB1 mutants. Full length YB1 and YB1 Δ 5 partially rescued the %cytotoxicity to a level similar to that in parental 293T control cells. The cytotoxic phenotype was further rescued by the stable expression of YB1 Δ 2- Δ 4, in fact conferring resistance to chloroquine-induced cell death. However, YB1 Δ 1 (which showed some disparity between chloroquine treatment conditions) and YB1 Δ 6 were unable to rescue the cytotoxic phenotype at all, with %cytotoxicity near comparable to levels exhibited by YB1 knockout cells after chloroquine treatment. It was also appreciated that the rescue of cell cytotoxicity was not hugely

dependent on the subcellular localisation of YB1 truncated mutants. This was inferred due to the fact that rescue of the phenotype was achieved with YB1 Δ 2, YB1 Δ 4 (showing cytoplasmic and nuclear localisations), YB1 Δ 3 (cytoplasmic), and YB1 Δ 5 (nuclear, with nucleolar predominance). Therefore, the resistance to chloroquine-induced cytotoxicity could be induced by a range of functions relevant to the expression of the respective YB1 domains. Resistant phenotypes were particularly established by an intact CSD and minimally CTD containing truncation mutants. The CSD demonstrated some particular importance to the resistance to chloroquine-induced cell cytotoxicity. This was an interesting observation because murine YB1 and MYS4 (murine Y-Box protein 1 paralogue; Ybx2) are essential in embryonic development (Lu *et al.*, 2005, 2006; Uchiumi *et al.*, 2006), with significant homology shared mainly between their CSD. YB1^{-/-} mice showed embryonic lethality after early stage embryonic development (Yalkinoglu *et al.*, 1988; Lu *et al.*, 2005; Uchiumi *et al.*, 2006), MYS4^{-/-} mice exhibited increased spermatocyte apoptosis, and double YB1^{-/-} and MYS4^{-/-} was completely incompatible even in early embryonic development (Lu *et al.*, 2006). Therefore, an association between the CSD and developmental viability is certainly acknowledged. This conserved function by murine YB1 and MYS4 may extend *in vitro* to alternative cell lines and their orthologues. Essentially, the protection/resistance to chloroquine-induced cell cytotoxicity may be a CSD-dependent mechanism that has yet to be elucidated in full.

Additionally, overexpression of YB1's CTD has been associated with suppression of cell proliferation and even tumour progression (Shi *et al.*, 2016; Wang *et al.*, 2016c). Wang and colleagues (2016c) found that overexpression of the CTD (125-220aa) correlated with increased apoptotic Bax and caspase 3 expression, and reduced anti-apoptotic Bcl-2 expression in EA.hy926 cells. Similarly, YB1's A/P domain has been previously found associated with cell cycle arrest at the G2/M phase and apoptosis (Khandelwal *et al.*, 2009). Cell cycle arrest and apoptosis was recorded by expression of as little as the first 77aa of YB1 protein (Khandelwal *et al.*, 2009). Particularly, the first 26aa of YB1's A/P domain was characterised as the major death domain by sequestering cyclin D1 in the cytoplasm (Khandelwal *et al.*, 2009). With this in mind, and the fact that truncation mutants YB1 Δ 1- Δ 3 harboured the A/P domain and CSD, the protective function of the CSD by the current study was perhaps further exemplified. However, given the above and that phenotype rescue was also partially achieved by YB1 Δ 5 (truncated mutant that encoded only the CTD), it could be reasoned that all three domains of wildtype YB1 may interplay with one another in a complex manner to confer

protection against chloroquine, as seen in parental 293T. For example, the CSD may exert particular functions that are either sequestered away or negatively regulated by the A/P domain and/or the CTD (concerning the sequences after the 20S proteasome cleavage site). This could lead to no rescue or partial rescue of cell cytotoxicity induced by chloroquine, as opposed to resistance shown by YB1's CSD alone.

It is noteworthy that the stable expression of YB1FL, YB1Δ1, YB1Δ2, YB1Δ5, and YB1Δ6 did not confer an additional protective phenotype against cytotoxicity exhibited by the PEI_{max}-mediated transfection. The YB1 knockout cells consistently demonstrated a comparable cytotoxicity profile to parental 293T. This further supports that an alternative pathway or mechanism involving chloroquine, but independent of PEI, is responsible for an enhanced cytotoxic phenotype in YB1 knockout cells to chloroquine reagent. PEI has been previously associated with cell cytotoxicity (Boussif *et al.*, 1995), and the mechanism by which this occurs has been investigated. Although, most studies explored the impact branched PEI has on cell cytotoxicity, linear PEI was used throughout the current study. It is also acknowledged that a more significant induction of cell death and cytotoxicity was reported by branched PEI compared to linear PEI (Kafil and Omid, 2011). Nevertheless, branched PEI has been shown to induce necrotic cell death and plasma membrane damage (Florea *et al.*, 2002; Moghimi *et al.*, 2005), and compromise the integrity of mitochondrial membrane potential by facilitating the release of cytochrome C (Moghimi *et al.*, 2005), proton leakage and inhibition of the electron transport chain (Hall *et al.*, 2013), thus potentiating in cell cytotoxicity and cell death via apoptosis. Although, Hall *et al.* (2013) do not directly correlate the inhibition of the electron transport chain and proton leakage with cell cytotoxicity in their studies, Moghimi *et al.* (2005) do correlate the increase in cytochrome C activity with cell cytotoxicity, but at very high concentrations of branched PEI. In the present study, approximately 10μg/mL of linear PEI was used to transfect cell lines – a concentration at which showed significantly less cytochrome C activity after 24h of treatment with free branched PEI (Moghimi *et al.*, 2005). YB1 has also been found to negatively regulate mitochondrial oxidative phosphorylation by suppressing the translation of associated mRNAs (Matsumoto *et al.*, 2012). Given this and the fact that YB1 knockout did not correlate with an enhanced cytotoxic response to PEI or PEI-mediated transfection compared to parental 293T, it could be surmised that the contribution PEI has on the mitochondrial proton leakage and inhibition of the electron transport chain is insignificant to the cytotoxic profile presented – at least at the working concentration of PEI used. Also, the

proposed mechanism in which PEI mediates cytotoxicity is, to date, independent of p53-associated signalling.

7.9 Limitations of applying CRISPR/Cas9 genome editing

The Sf9 cell line has shown promising capacity for heterologous protein expression by infection with recombinant baculoviruses (Graber *et al.*, 1992; Miranda *et al.*, 1997; Usami *et al.*, 2011; Wilde *et al.*, 2014). This application was then harnessed for scalable rAAV vector production by using recombinant baculovirus technology (Urabe *et al.*, 2002; Urabe *et al.*, 2006). And since has proven as a clinically relevant producer cell model for rAAV vector production (Carpentier *et al.*, 2012). CRISPR/Cas9 genome editing was performed on Sf9 cells, targeting the putatively identified *SFYB* gene that encodes the SfYB protein. The approach in which we mediated CRISPR/Cas9 genome editing was explored in the present study – plasmid transfection or recombinant Cas9 and synthesised sgRNAs. Transfection of Sf9 cells with pCRISPR plasmids was complicated by almost negligible GFP expression 72h post-transfection, despite having had optimised plasmid transfection in Sf9 cells using the Cellfectin™ II Reagent and pEx-1.eGFP reporter plasmid. The efficiency of transfection does not correlate with genome editing; however, SpCas9 is co-expressed with GFP reporter and the poor transfection efficiency with pCRISPR plasmids may infer an issue in SpCas9-2A-GFP expression. This was problematic because SpCas9 is the main effector molecule that enzymatically induces DSBs and promotes gene editing (Gasiunas *et al.*, 2012; Wei *et al.*, 2015b). It was noted that SpCas9-2A-GFP expression was under CMV promoter control, which was in contrast to the IE1 promoter and hr5 enhancer for pEx-1.eGFP. The IE1 promoter is a baculovirus immediate-early gene promoter useful for strong heterologous gene expression in insect cells (Masumoto *et al.*, 2012). The CMV promoter is commonly used for the expression of high levels of recombinant protein in mammalian cell lines (Wang *et al.*, 2017). Despite this, the CMV promoter was previously reported to drive transgene expression in Sf9 cells, either by baculovirus infection or plasmid transfection (Li *et al.*, 2014). Albeit, this was less efficient compared to the baculovirus-specific polyhedrin promoter (Li *et al.*, 2014). Additionally, the CMV-p10 fusion promoter was also agreeable for protein expression in Sf21 cells, post-infection (Mansouri *et al.*, 2016). However, the contribution of the CMV promoter in transgene expression was not differentiated from p10 by Mansouri *et al.* (2016). Alternative reports infer that the CMV promoter is weakly active in Sf9 cells, as indicated by near negligible luciferase activity (He *et al.*, 2008), and thus in agreement with poor GFP expression frequencies reported in the current study.

Therefore, the weak expression of the SpCas9-2A-OFP using GeneArt™ CRISPR nuclease vectors was considered inadequate for CRISPR/Cas9 genome editing of Sf9 cells. Alternatively, transfection of Sf9 cells with sgRNAs generated by IVT and recombinant Cas9 for *SFYB* targeting was instead implemented. This off-sets the need for incompatible promoters to drive the expression of the Cas9 nuclease effector in choice cells, without compromising gene editing frequencies as previously reported (Liang *et al.*, 2015; Yu *et al.*, 2016; Seki and Rutz, 2018). It was previously acknowledged that transfection of Sf9 cells (and insect cell lines in general) with naked plasmid DNA is typically associated with poor to moderate transfection capacities – ranging between 30-50% (Maeda *et al.*, 2005; Ogay *et al.*, 2006). We do find further optimised transfection efficiencies using CRISPRmax™ Transfection Reagent and opti-MEM™ as the diluent in the present study. High and efficient transfection rate of nearly 100% was achieved with as little as 0.5µg of plasmid per 4.5×10^5 Sf9 cells. Therefore, it was inferred that highly efficient transfection of IVT sgRNAs and recombinant Cas9 RNP complex could similarly be achieved to enhance the genome editing rates of the *SFYB* gene.

Furthermore, the SfYB protein was identified as a Y-Box protein homologue, and the expression of which was consolidated by using SfYB-specific antiserum, raised specifically against rSfYB(His)₁₀ protein. Further to this, we were able to BLAST for the gene sequence encoding the SfYB protein (*SFYB*) from the previously assembled contig (OE0A01010394.1), and annotated intron/exon sequences based on sequence homology to *S. litura*'s putatively identified *SLYB* gene sequence (NC_036217.1; LOC111360813). This was an essential requirement for designing gRNAs to target the putative *SFYB* gene for targeted knockout using CRISPR/Cas9 technology. CRISPR/Cas9 genome editing requires a gRNA – a crRNA-tracrRNA (otherwise synthesised as the sgRNA for *in vitro* applications) species – that is specific to the sequence of interest in order to facilitate Cas9 targeting and targeted nuclease activity (Jinek *et al.*, 2012). Additionally, the specificity of gRNAs promotes on-target efficiency and reduces off-target effects (Fu *et al.*, 2014; Dang *et al.*, 2015; Xu *et al.*, 2017). The gRNAs designed in the present study were restricted to exon-specific sequences given the putative nature of the mapped gene sequence. This was also appreciative of the fact that the genomic origin of target sites directly dictates the efficiency of sgRNA-Cas9 targeting, with intron-specific sgRNAs performing worst (Labuhn *et al.*, 2018).

Initial screens for knockout or partial knockout of SfYB protein expression was performed using αYB1 raised against human YB1 peptides. The lack of commercially available αSfYB

restricted our screening capabilities by Western blotting to α YB1. This was due to the lack of complete and well annotated proteomics data for *S. frugiperda*, and instead proteins are usually identified by aa sequence homology and protein annotations from other insect species. Few exceptions of endogenous proteins for *S. frugiperda* have been successfully cloned and characterised (Ahmad *et al.*, 1997; Silva *et al.*, 2013). Quantitative proteomics between control and baculovirus infection has also suggested host-baculovirus interactions (Carinhas *et al.*, 2011; Yu *et al.*, 2015). However, the Y-Box protein homologue for *S. frugiperda* has not been previously described until our current studies.

Given that the homology between YB1 and SfYB is not 100%, we opted to express and purify recombinant SfYB protein to raise SfYB-specific antiserum for complete validation of gene disruption. Interestingly, initial Western blotting demonstrated a disparity in SfYB-signal and expression between using α YB1 and the SfYB-specific antiserum after sgRNAsf1:Cas9 RNP transfection. In fact, SfYB knockout, at least by the limits of detection by Western blotting, was achieved using the SfYB-specific antiserum. This disparity was likely a result of poor cross-reactivity of the α YB1 given the reduced aa homology between human YB1 and SfYB; and was instead corrected by the more specific nature of α SfYB. We thus demonstrated the importance of target-specific primary antibodies for the screening of CRISPR/Cas9 genome edited cell lines.

We also emphasise the importance of knowledge of genomic DNA sequences and verified annotated coding sequences for CRISPR/Cas9 functionality. This goes without saying given that precise Cas9-mediated DNA targeting is sequence- and PAM- dependent (Anders *et al.*, 2014; Kleinstiver *et al.*, 2015). Although, advancements in the CRISPR/Cas9 genome editing has shown that genomic DNA-specific gRNA-free genome editing has been developed by Zhao and colleagues (2017). This was performed in an *E. coli* system, but was dependent on HR of a universal DNA sequence (N20PAM) for Cas9-targeting into the *E. coli* chromosome. Therefore, the process still required precise knowledge of DNA sequences. Initial attempts to identify the *SFYB* gene, and in turn design sequence-specific gRNAs, was very restricted by the limited availability of reference genomes for *S. frugiperda* (Kakumani *et al.*, 2014). This incentivised the designing gRNAs based on sequence homology to mapped exons using *BYB* and *SLYB* as genomic scaffolds. The gRNAs designed in the current study were therefore solely exonic in their genomic origin to improve targeting and efficiency of CRISPR/Cas9 genome editing

(Labuhn *et al.*, 2018). The complete *S. frugiperda* reference genome assembly has recently been established as of late 2017 (Gouin *et al.*, 2017; Nandakumar *et al.*, 2017).

Identification of off-target cleavage also requires the utilisation of complete reference genomes for screening experimentally or *in silico* (Kim *et al.*, 2015; Tsai *et al.*, 2015). Off-target cleavage may potentially confound the interpretations of observations and experimental results. This is further complicated in practice with the genetic variation existing in the population of a given species (Genomes Project *et al.*, 2012; Genomes Project *et al.*, 2015). This includes *S. frugiperda*, which also exhibits significant genetic variation between its rice- and corn- eating sympatric strains (Gouin *et al.*, 2017). This is less an issue *in vitro* with experimental cell lines, which are generally clonally derived and reference genomes are typically correlating. However, there remains a probability that genetic variation – in the form of polymorphisms and indels – could result in altered on- or off- target sites (Lessard *et al.*, 2017). Ultimately, and especially concerning human clinical trials and patients, this could predispose subjects to undesired and even deleterious effects when using CRISPR/Cas9 technology. The lack of complete and well annotated genomic DNA data likely limits the use of CRISPR/Cas9 genome editing across a number of cell lines. Therefore, the full utilisation of CRISPR/Cas9 genome editing is restricted by the extent and depth of available reference genome data – both for on- and off- targets.

We previously reported that single cell cloning of pCRISPR-transfected 293T cells was beneficial for the establishment of single cell clones. The absolute knockout of YB1 and mutations in its encoded *YBX1* gene was confirmed; and the approach promoted the establishment of YB1 knockout cell lines that were homozygous for the CRISPR/Cas9-mediated mutation amongst all *YBX1* alleles. However, Sf9 cells – either non-transfected or transfected with CRISPR RNP complex, failed to grow as a monoclonal cell population. This was even in spite of the fact that Sf9 cells are clonal cell derivatives of the Sf21 cell line (Vaughn *et al.*, 1977). Monoclonal cell populations failed to grow and expand despite the variations of media composition used in the current study, including conditioned media. The addition of conditioned media has been shown to stimulate Sf9 cell proliferation in culture (Calles *et al.*, 2006), although the effect on single cells has not been elucidated. It was noted that Sf9 cells showed cell density-dependent growth characteristics, which is a general characterisation of both Sf9 and Sf21 cell lines. Therefore, the establishment of single-cell clones by limiting dilution was perhaps too difficult to achieve for Sf9 cells. This also prevented the

establishment of stable SfYB-disrupted or knockout cell lines using CRISPR/Cas9 genome editing.

7.10 YB1 and SfYB knockout cell lines do not enhance rAAV vector titres

It was previously reported that YB1's association with rAAV vectors by Satkunanathan *et al.* (2014) may implicate the endogenous host cell protein (YB1) with rAAV processing. This was especially promising given that YB1 knockdown using shRNA in 293T cells correlated with an enhanced rAAV2 and rAAV8 producer cell line quality, with up to 45-fold more rAAV2 vector genome titres calculated compared to control cell lines (Satkunanathan *et al.*, 2014). Although, the transient nature of this enhanced producer cell line quality limited the applicability of the novel cell line for further research. For example the transient enhanced producer cell line quality meant a long-term, stable solution to generate rAAV vector quantities that meet the demands of clinical studies had not been achieved.

To supplement this novel approach to enhance rAAV vector production, we endeavoured to utilise CRISPR/Cas9 genome editing to manipulate host cellular factors, particularly by targeting YB1 expression for knockout. Initial attempts to triple transfect 293T and YB1 knockout cells for rAAV2 vector production using the Calcium Phosphate precipitation method was hindered experimentally by the acute and heightened sensitivity to chloroquine. However, despite overcoming the unprecedented complication and limitation of the YB1 knockout cell lines to the aforementioned selected transfection modality (Calcium Phosphate precipitation method supplemented with chloroquine), an enhanced rAAV2 vector producer cell line was not identified. In fact, the YB1 knockout cell line generated an equivalent quantity of rAAV2 vector genome titres when compared to 293T control. Similarly, rAAV2 vector genome titres produced from YB1 knockout and mutant YB1 Δ 1- Δ 6-expressing cells were comparable to that from 293T control. The expression dynamics of Rep and Vp1-3 proteins was considered overall unaltered by YB1 knockout. And the localisation of AAV2 intact particle, Rep and Cap proteins was generally unchanged by YB1 knockout phenotype or the stable expression of YB1 Δ 1- Δ 6. Therefore, the YB1 knockout phenotype in 293T did not confer an enhanced producer cell line.

It was noted that rAAV2 vector production was mediated by triple transfection using PEI_{max} transfection reagent instead of Calcium Phosphate precipitation method, as was used in the previous study by Satkunanathan *et al.* (2014). However, PEI-mediated transfections function

rather analogous to Calcium Phosphate transfection mechanism, in that DNA:PEI complexes are receptive to internalisation by endocytosis (Utsuno and Uludag, 2010). Furthermore, analogous to chloroquine, PEI protects input DNA by buffering the acidic pH of endosomes (Boussif *et al.*, 1995; Sonawane *et al.*, 2003). The cytotoxic potential and differences exhibited on cultured cell lines by PEI or Calcium Phosphate precipitation methods of transfection were screened and thus controlled for by LDH cytotoxicity assays in the current study. However, it remains unclear why an overall enhancement in rAAV2 vector genome titres was not a feature of YB1 knockout cells. Although, the mass ratio of the three AAV plasmids used to triple transfect 293T and YB1 knockout cells in the current studies (1:3:1 mass ratio of pAAV2-hrGFP, pHelper, and pAAV2/2-RC) was in some contrast to the mass ratio of plasmids used by Satkunanathan *et al.* (2014) – 1:1:3 mass ratio of pAAV2-hrGFP, pHelper, and pAAV2/2-RC, respectively. It is therefore noted that on a per tissue culture plate basis, a larger quantity of each and total plasmid was used by the study performed by Satkunanathan *et al.* (2014) – 37.5µg total plasmid using Calcium Phosphate precipitation method compared to 23µg total plasmid using PEI_{max}. Although, it is difficult to attribute the lack of enhanced rAAV vector titres from YB1 knockout cells by the difference in plasmid quantity and ratio used, mainly because different transfection methodologies were used to produce rAAV2 vectors.

Furthermore, YB1 has demonstrated in previous reports to encourage the production dynamics of a number of viruses and viral vectors, including HIV1 (Mu *et al.*, 2013), recombinant MLV (Li *et al.*, 2012), Influenza A virus (Kawaguchi *et al.*, 2012). But negatively regulates DV by functioning in an antiviral manner to repress the translation of viral proteins or RNA replication (Paranjape and Harris, 2007). Here, YB1 knockout MEF lines were associated with significantly enhanced DV replication, production and viral protein expression (Paranjape and Harris, 2007). Associations between YB1 and Adenovirus have also been investigated, and Adenovirus viral factors required for rAAV2 vector production by transient transfection of 293T cells include helper factors encoded by *E1*, *E2A*, *E4* and *VA* genes (Matsushita *et al.*, 1998; Xiao *et al.*, 1998). These are supplied either *in trans* by the pHelper plasmid or stably expressed by 293T cells. YB1 has been reported to impact on the splicing characteristics of Adenovirus E1A pre-mRNA (Chansky *et al.*, 2001; Raffetseder *et al.*, 2003), the gene activity of E2's late promoter (Holm *et al.*, 2002). The close association between YB1 and Adenovirus E1B protein and its co-targeting to viral inclusion bodies presumably thought to promote Adenovirus replication (Holm *et al.*, 2002). However, when concerning wildtype AAV or rAAV vectors, the functions of *E1* and *E2* gene products are essential for AAV

processing, given that these elements are some of a few minimum essential helper factors necessary for rAAV vector production (Matsushita *et al.*, 1998; Xiao *et al.*, 1998). The function of E1A promotes up-regulation of AAV promoters – p5 and p19 (Tratschin *et al.*, 1984), E1B promotes nuclear export of AAV late viral mRNA (Krätzer *et al.*, 2000; Blanchette *et al.*, 2008), and E2A functions to promote AAV replication (Ward *et al.*, 1998). Especially, when considering the functions that E1A, E1B and E2A have in context with AAV biology, and the impact YB1 has on the processing of these proteins, the complete loss of YB1 expression in YB1 knockout cell lines generated in the present study may not be completely conducive for enhanced rAAV2 vector production. However, the relevance of YB1 to these helper proteins has only been identified in context to Adenovirus gene elements or infection, and therefore is not strictly AAV-associated.

7.11 The role of YB1/SfYB protein and its association with rAAV vectors

We demonstrated that AAV Rep and Cap proteins were expressed comparably between control 293T or Sf9 and the corresponding CRISPR/Cas9 genome edited cell lines. This was in contrast to the improved Rep protein expression and vector genome titres by the YB1 knockdown cell line, although derived from 293T, using shRNA technology (Satkunanathan *et al.*, 2014). It was noted that the expression of AAV Rep and Cap proteins using recombinant baculoviruses are not under their wildtype AAV p5, p19 and p40 promoters as we see in AAV plasmids (Matsushita *et al.*, 1998; Xiao *et al.*, 1998). Instead, AAV2 *rep* ORF is under promoter control of polh and the serotype-specific *cap* ORF is under p10 promoter control (Smith *et al.*, 2009). Therefore, the expressions of these proteins are entirely dictated by the host-cell response to baculovirus infection. Whether or not the Y-Box protein homologue, SfYB, is able to influence the activity of baculovirus-derived promoters has not been previously identified. However, human YB1 is a broadly functioning host-cell factor that is widely implicated in regulating exogenous viral promoters. This includes Adenovirus *E2* late promoter, of which binding promotes the *E2* gene expression (Holm *et al.*, 2002). Additional examples include the JC virus late gene promoter (Kerr *et al.*, 1994; Chen and Khalili, 1995), and HIV1's LTR promoter (Sawaya *et al.*, 1998), which exhibit enhanced transcriptional activities upon YB1 binding.

However, the *E2* gene helper function is omitted from rAAV vector production using baculovirus infection. The helper function is substituted by the productive baculovirus infection instead to drive p10 and polh promoter activities. Therefore, the Adenovirus helper

genes are absent from rAAV vector processing. This includes the *E2* gene, in particular the E2A product, which has been recognised to complement AAV viral DNA replication (Ward *et al.*, 1998) and viral transcription (Chang and Shenk, 1990). E2A was shown to promote elongation of replicating AAV genomes (Ward *et al.*, 1998). Vector production in Sf9 is mediated by the natural assembly of AAV capsids and replicating AAV ssDNA genomes as per AAV2 Rep proteins and host cellular machineries (Ruffing *et al.*, 1992). E2A protein has been shown to localise with AAV viral DNA (Pombo *et al.*, 1994), but also associates with Rep78/68 and AAV *ITRs* (Stracker *et al.*, 2003). The purpose of which was suggested to enhance protein binding to promote replication and DNA nicking by terminal resolution (Stracker *et al.*, 2003). Although, further interplay between YB1 and E2A beyond its binding to the *E2* late promoter (Holm *et al.*, 2002) has yet to be identified, if any. We have identified relatively specific YB1 binding to AAV viral DNA – especially to the *ITRs*. Therefore YB1, and by extension SfYB, may compete for Rep78/68-E2A binding to AAV genomes, ultimately affecting the processivity. The optimal processivity of rAAV vectors in Sf9 cell lines or its derivatives may be restricted by the absence of Adenovirus E2A protein, and perhaps the SfYB knockout phenotype could not be fully taken advantage of.

We also appreciate that human YB1 is implicated in the control of endogenous gene expression, mRNA transcription, and protein translation (Didier *et al.*, 1988; Evdokimova *et al.*, 2001; Skabkin *et al.*, 2001; Stickeler *et al.*, 2001; Hartmuth *et al.*, 2002; Pisarev *et al.*, 2002; Higashi *et al.*, 2003; Lasham *et al.*, 2003; Nekrasov *et al.*, 2003; Homer *et al.*, 2005; Deckert *et al.*, 2006; Kotake *et al.*, 2013). Interestingly, YB1 generally exhibit functions to inhibit translation or stabilise mRNA. YB1 is a major protein component of non-polyribosome-bound mRNPs (Kumar and Pederson, 1975; Minich *et al.*, 1993), and regulates the assembly of the 40S pre-initiation complex to the translation initiation codon of mRNA (Pisarev *et al.*, 2002). With low YB1:mRNA ratio, YB1 binds to mRNA as monomer units along the length of mRNA using its CSD and CTD, but homo-oligomeric YB1 complexes associate with mRNA instead when YB1:mRNA ratio is high (Davydova *et al.*, 1997; Pisarev *et al.*, 2002). Homo-oligomeric YB1 complexes to mRNA may impede the association of necessary translation factors because mRNA-YB1 complexes may be packaged in such a way that its termini are thought to be buried, ultimately preventing the translation machinery from assembling (Skabkin *et al.*, 2004). In contrast to this, YB1-depleted lysates were found to be almost completely deficient of protein translation of the β -globin mRNA (approximately 90% inhibition with anti-YB1-mediated YB1 depletion in rabbit reticulocytes) using a cell-free *in vitro* protein translation

system (Evdokimova *et al.*, 1998). Furthermore, YB1-depletion was also associated with the accelerated decay of mRNAs, again using rabbit reticulocyte-derived lysates and cell-free *in vitro* systems (Evdokimova *et al.*, 2001). The complete depletion of YB1 or its overexpression relative to mRNA levels was found not to be conducive for *in vitro* protein translation at the initiation stage; and low YB1:mRNA ratios was optimal for efficient protein translation. Therefore, the YB1/SfYB knockout phenotypes displayed by YB1/SfYB-disrupted cell lines were perhaps not conducive for the improved expression of AAV Rep or Cap proteins, in the current study. This implies that gene manipulation/knockout by CRISPR/Cas9 genome editing may not be wholly appropriate given the probable nature and complexity of human YB1 and SfYB's function *in vitro*.

It could be reasoned that the lack of enhanced AAV2 *rep* expression in the current studies was a product of compromised protein translation machinery. YB1-depletion using α YB1 (Evdokimova *et al.*, 1998; Evdokimova *et al.*, 2001), perhaps most closely resembled the YB1/SfYB knockout phenotypes generated by CRISPR/Cas9 genome editing system. However, the current studies focused on knockout of YB1/SfYB expression in 293T or Sf9 cells, which have human or insect origins, respectively. Nonetheless, the cell-free *in vitro* systems identified the inhibitory effect of YB1, but was limited to using β -globin or luciferase mRNA as their substrate, only. We previously have shown that the expression of a number of endogenous proteins (SOX13, SDF4, PANK4, and GAPDH) was relatively unchanged between parental 293T and YB1 knockout cell lines when testing for CRISPR/Cas9 off-targeting. We thus simultaneously tested the effect of YB1 knockout phenotype on some endogenous protein expressions. The impact YB1 depletion has on mRNA decay and/or translation inhibition (Evdokimova *et al.*, 1998; Evdokimova *et al.*, 2001) was rather contested by the YB1 knockout cell lines. It is appreciated that the mRNA decay and translation initiation of AAV2 Rep mRNA and even for Vp1-3 were not explored in the current study, and may be compromised by the YB1 knockout phenotype.

If YB1-depletion is any indicator to the potential impact the YB1 knockout phenotype has on rAAV2 Rep and Cap expression (either on mRNA or protein translation), then this could explain the comparable rAAV2 vector genome titres between control and genome edited cell lines. Whereby, the lack of enhancement of AAV2 *rep* expression in YB1 knockout cell line would be unable to promote rAAV genome replication at an enhanced rate for packaging and ultimately enhanced rAAV vector titres. This was perhaps the molecular milieu attained in YB1

knockdown cell lines produced by Satkunanathan *et al.* (2014), which was permissive for enhanced AAV2 *rep* expression, and may have dictated the enhanced vector genome titres calculated. This adds another complication of targeting YB1 for CRISPR/Cas9 genome editing that may confer a less than conducive cell line for rAAV vector production, after having targeted an extremely multifunctional protein.

Regardless of the lack of effect identified in YB1/SfYB knockout cell lines generated in the current studies, YB1 has demonstrated a clear role in rAAV vector processing, as determined by Satkunanathan *et al.* (2014). The identification of wildtype YB1 in rAAV vector particles by liquid chromatography-mass spectrometry/mass spectrometry (Satkunanathan *et al.*, 2014) suggests that it is at the very least a cell-derived impurity that is packaged into rAAV. ICC and confocal microscopy was used to trace the localisations of rAAV2 intact particle, AAV2 Rep and Cap proteins, relative to wildtype YB1 (including truncated YB1 mutants and YB1 knockout phenotype). It was generally considered that the baseline localisations of wildtype YB1 and YB1 mutants were unchanged between non-transfected and triple-transfections. Also, the localisations of intact rAAV2 vector particle, AAV2 Rep and Cap proteins were unchanged between control and knockout cell lines, suggesting that YB1/SfYB does not possess an obvious role in trafficking rAAV2 vectors during its production. However, we have identified a rather distinct colocalisation between YB1 Δ 5 (encompassing wildtype YB1's CTD) and rAAV2 intact particle and AAV2 Cap within the nucleus. This localisation was more specifically localised within the nucleolar compartment. Therefore, there seems to be an association between YB1's CTD and rAAV2 vector during the production process.

AAV assembly of progeny virions or rAAV vectors has been reported to be concentrated in the nucleolar regions (Wistuba *et al.*, 1997), a process which is mediated by the AAV2's AAP scaffolding protein (Sonntag *et al.*, 2010; Naumer *et al.*, 2012; Earley *et al.*, 2015). Therefore, the assembly of rAAV2 capsids may be one potential point at which wildtype YB1 is packaged into rAAV2 vectors as a cell-derived impurity. However, the relevance for YB1's association with rAAV2 assembly within the nucleolus via the CTD is unclear given no significant effect on rAAV2 vector genome titres was observed from +YB1 Δ 5 cell line, relative to 293T control. On the other hand, YB1 interaction with AAV *ITR* sequence was identified in the current study. The AAV *ITRs* are particularly important in the encapsidation and packaging of AAV genomes into preformed capsids (Wang *et al.*, 1996; Dubielzig *et al.*, 1999; Bleker *et al.*, 2006).

Wildtype YB1 was also found to associate with AAV and Adenovirus DNA sequences by DNA affinity-YB1 pulldown studies. YB1 was previously identified to bind to the Y/CCAAT box (also known as the Y box) motif – 5'-CTGGATTGG C/T C/T AA-3' (Didier *et al.*, 1988). However, its ability to bind to dsDNA and ssDNA species was not confined to the Y box (Izumi *et al.*, 2001; Zasedateleva *et al.*, 2002; Kljashtorny *et al.*, 2015), and preferential binding was observed for ss GGGG motifs (Zasedateleva *et al.*, 2002). In the present study we identified putative YB1 binding motifs in the AAV2 *ITRs*, within the early coding sequences of AAV2 *rep*, and upstream of Adenovirus *E2a* gene sequence. Of particular interest was the AAV2 *ITR*, of which putative YB1 binding motif (5'-GGGGTT-3') was found present in the pAAV2-MCS plasmid (and by extension any AAV2 transfer vector used for rAAV vector production), and in the -sense strand of the rAAV vector genomes. Importantly, the *ITR*-containing AAV2 transfer vector is the initial copy used as a template for the replication of rAAV vector genomes.

Despite the transient nature of input plasmid during the course of *in vitro* transfection, especially without antibiotic selection, the ds *ITR*-containing plasmid still functions as target for potential binding for YB1. YB1 was captured and pulled down using relatively long 5'-desthiobiotin-labelled capture probes (dsDNA), with the putative 5'-GGGGTT-3' YB1 binding motif centrally positioned. Similar was demonstrated for the *rep* and *E2a* capture probes. However, the specificity of these observed binding and pulldowns was questionable given the control capture probes (especially the right *ITR* flanking control probe) also demonstrated YB1 pulldown. These control probes were initially screened for putative YB1 binding motifs as per Zasedateleva *et al.* (2002), of which only the left *ITR* flanking capture probe harboured. The right *ITR* flanking control capture probe was still able to significantly bind and pulldown YB1 to comparable levels as to the substrates in question. Upon closer examination, the right *ITR* flanking control capture probe harbours a single 5'-GGCGGG-3' (GC-box) motif, that has been reported to show high specific binding to YB1 (Shi *et al.*, 2012; Shi *et al.*, 2013). Although, Shi *et al.* (2013) report using a quadruplet 5'-GGCGGG-3' oligo in their YB1 pulldown studies, the highly specific binding of YB1 to the GC-box found in the right *ITR* flanking control probe could explain the YB1 pulldown that could not be outcompeted by the poly(dI-dC) competitor. It is also worth noting that a single copy of the GC-box motif was identified in the *ITR* (putative 5'-GGGGTT-3' containing) and *E2a* (putative 5'-CCTCCT-3' containing) capture probes, but absent in the *rep* (putative 5'-GGGGTT-3' containing) and left *ITR* flanking control capture probes. Therefore the binding and pulldown by the *ITR* and *E2a* capture probes may hold merit given the *rep* capture probe, which is without the GC-box, was still able to bind and pulldown YB1

similarly. The specificity of the binding was finally verified for the *ITR* putative YB1 binding motif by using shorter ds and ss oligos that were void of the GC-box, and using an irrelevant *Luciferase* transgene sequence as a control capture probe.

It is worth addressing the fact that the rAAV vector genome's left *ITR* of the +sense strand harbours the putative 5'-GGGGTT-3' YB1 binding motif, which is in turn positioned 78bp upstream of the transgene's CMV enhancer/promoter sequence. Additionally, the right *ITR* of the +sense strand possesses the *ITR*-derived 3'-hydroxyl group for unidirectional DNA replication, which upon the generation of a dsDNA intermediate, potentiates in the putative ds 5'-GGGGTT-3' YB1 binding motif. Therefore, given YB1 is packaged into rAAV vectors (Satkunanathan *et al.*, 2014), and YB1 binding to these sequences is very much probable, with the successful YB1 pulldowns achieved by the ss*ITR* capture probes in the present study, it could be reasoned that the packaging of YB1 is mediated at this very point of rAAV vector processing – rAAV genome packaging into preformed capsids. Alternatively, the interaction could suggest a means in which YB1 regulates rAAV genome packaging. The encapsidation process of AAV genome occurs in a seemingly polar manner – the 3'-ends are translocated into preformed capsids first via Rep40/52's 3'-5' helicase/ATPase activity (King *et al.*, 2001). Rep78/68 protein complex with the AAV2 *ITR* sequences is thought to help dock genomes to preformed capsids (Dubielzig *et al.*, 1999; Bleker *et al.*, 2006), especially is close association with the pores at the five-fold symmetry axes (Bleker *et al.*, 2005). Not only this, the *ITR* D-sequences of the AAV2 *ITRs*, which harbours the 5'-GGGGTT-3' putative YB1 binding motif, is essential for AAV genome encapsidation, as its deletion correlates with inefficient packaging of the AAV genome into preformed capsids (Wang *et al.*, 1996). Therefore, with YB1's relatively specific and high affinity binding to *ITR* sequences demonstrated in the present study, it could be reasoned that Rep proteins are in direct competition with YB1 for binding to the *ITR*. Although, AAV2 Rep78 was also found to bind to YB1 by co-immunoprecipitation experiments (Nash *et al.*, 2009), the functional relationship between this interaction was not elucidated. Regardless, with this competition in binding to the AAV2 *ITR* comes a potential interplay that regulates or dictates the encapsidation of rAAV vector genomes.

7.12 Significance of CRISPR/Cas9 genome editing for AAV gene therapy development

Utilisation of CRISPR/Cas9 genome editing has begun to revolutionise human gene therapy, potentially giving rise to an entirely new class of therapeutics. Harnessing the genome editing technology has already shown promising potential as a gene therapy strategy for the

treatment of HIV1 (Kaminski *et al.*, 2016). In particular, a number of studies have successfully used a combination of AAV and CRISPR/Cas9 genome editing, wherein the vectors' cell/tissue tropism for transduction was exploited. AAV-CRISPR systems developed for targeted correction of haemophilia B mice models have shown curative potential for the disease phenotype (Guan *et al.*, 2016; Ohmori *et al.*, 2017). Additional *in vivo* use of CRISPR/Cas9 has generally exemplified promising correction of disease phenotypes (Wu *et al.*, 2013; Guan *et al.*, 2016; Long *et al.*, 2016; Nelson *et al.*, 2016; Tabebordbar *et al.*, 2016; Huai *et al.*, 2017; Ohmori *et al.*, 2017)

Alternatively, harnessing the genome editing tool to improve or enhance a gene and/or cellular therapy product is a rather novel research concept. The idea to improve gene or cellular therapy products using CRISPR/Cas9 has been demonstrated with the engineering of CAR-T cells for endogenous TCR and/or PD-1 knockout. The ultimate goal is to improve the efficacy of the engineered T cells to function as tumour-specific T-cells (Blank *et al.*, 2006; van Loenen *et al.*, 2010; Wei *et al.*, 2013; Bunse *et al.*, 2014; Wu *et al.*, 2014). Therefore, the application of this concept could likely extend to alternative cell or gene therapy products.

A number of cell-intrinsic factors have evolved to function as host restriction factors for viral infection and propagation (Laguet *et al.*, 2012; Heusinger *et al.*, 2015). The advantage of these is to confer cellular resistance to naturally-occurring viruses in nature. However, such restriction factors can be hindrance to *in vitro* propagation of pseudotyped viral vectors (Mitchell *et al.*, 2014; Satkunanathan *et al.*, 2014). This especially includes rAAV vectors, which, for the most part are near comparable to their wildtype counterparts – an encapsidated ssDNA genome in a capsid shell. This is despite the introduction of *ITR*-flanked AAV transgene and minimal packaging and/or helper factors mediated by plasmid transfection or baculovirus infection. The relevance of AAV-host interactions directly or indirectly impacts on the ever present limitation of using rAAV vectors for gene therapy/transfer applications i.e. the constrained quantities produced by current systems. Therefore, gene regulation of characterised or putative restriction factors on rAAV biology with minimal changes in the genomic blueprint is an attractive means to establish cell lines for high titre rAAV vector output. Targeted and controlled modulation of host cell proteins that otherwise hinder rAAV vector processing could lead to optimal rAAV processing and yields. Clear benefits of establishing such cell lines could include the generating high doses of rAAV vector that are

required to overcome the inefficient transduction efficiencies *in vivo* (Pajusola *et al.*, 2002), and/or pre-existing immunity to AAV (Mingozzi *et al.*, 2013; Katz *et al.*, 2014).

7.13 Future perspectives

Genome engineering and gene expression manipulation of 293T or Sf9 using CRISPR/Cas9 did not confer an enhanced rAAV vector producer cell line. However, the work presented represents an alternative application of the technology for gene therapy product development. Despite the knockout of YB1 did not correlate with enhanced vector production, it may still be advantageous to explore the CRISPR/Cas9 genome editing and knockout of alternative cell-intrinsic factors that have been associated as AAV host-restriction factors. In turn, examining the effect of knockout of these factors in AAV processing could aid in elucidating their function(s) in this context. Of particular interest are cell-intrinsic factors which are unlikely to sensitise 293T cells to chloroquine-induced cell cytotoxicity. Examples of cell-intrinsic factors could include the promyelocytic leukaemia protein (PML), of which overexpression correlated with inhibition of AAV genome replication (Mitchell *et al.*, 2014). Its knockdown or knockout may correlate with enhanced AAV genome replication. Perhaps, further elucidating the exact role of cell-intrinsic factors may lead to a combination of targeted gene regulation and/or editing to promote optimal vector production and output.

Additionally, in order to further explore the utility of Sf9 cell line for rAAV vector production, the complete elucidation of host-virus interactions is warranted. Especially, identifying Sf9-derived cell-intrinsic factors that associate with rAAV vectors and processing is warranted, including the direct association between SfYB and AAV using liquid chromatography-mass spectrometry/mass spectrometry. Protein expression changes between mock baculovirus infection and AAV vector production – synonymous to research on the changes in Sf9 protein expression after baculovirus infection (Carinhas *et al.*, 2011; Yu *et al.*, 2015) – could infer probable host-virus interactions. Likewise could be performed for the *T. ni*-derived Ao38 cell line for rAAV vector production, which are a recommended cell line for enhanced recombinant protein expression compared to Sf9 or Sf21 (Wickham and Nemerow, 1993; Wilde *et al.*, 2014). The Ao38 cell line also has shown the propensity for rAAV vector production (Meghrou *et al.*, 2005), yet remains to become a clinically relevant producer cell line like its Sf9 counterpart.

The examination and elucidation of the complete functions of SfYB in the context of *S. frugiperda* (Sf9) physiology may be justified. Significant homology at the aa sequence level does not directly correlate with shared functions. This would require elucidating experimentally, but we show that SfYB is at the very least a dsDNA binding protein with little sequence specificity. It would also be advantageous to explore whether or not the disruption of SfYB may promote the production of recombinant baculovirus(es). This need not be exclusively performed using CRISPR/Cas9 genome editing; because of the possible translation of human YB1 depletion and its association with deficient protein translation (Evdokimova *et al.*, 1998). However, the potential downstream and commercial application of an enhanced baculovirus producer cell line or cell line exhibiting enhanced productivity are relevant and active research fields (Lai *et al.*, 2012; Gomez-Sebastian *et al.*, 2014; Lopez-Vidal *et al.*, 2015; Steele *et al.*, 2017). A key example of the latter is the two-fold improvement in recombinant protein expression in stable *Sf*-caspase-1 knockdown cell lines (Lai *et al.*, 2012). Regardless, elucidating the functional nature of YB1/SfYB binding to AAV2 *ITR* sequences is warranted to identify how exactly Y-Box proteins negatively regulate AAV vector production. *In vitro* transcription and translation studies of AAV *rep* or *cap* (with wildtype Adenovirus or baculovirus specific promoters) in the presence of recombinant YB1 and SfYB protein may clarify how these Y-Box proteins impact on these gene expressions.

Alternatively, it is appreciated that secreted YB1 protein is a mediator of the inflammatory response (Frye *et al.*, 2009; Hanssen *et al.*, 2013), and implicated in inflammatory diseases (van Roeyen *et al.*, 2005). Therefore, the packaging of 293T-derived YB1 (a cell derived impurity) into rAAV vector capsids potentiates in the mitogenic stimulation of the subject's immune response by exogenous YB1 post-administration. Therefore, the utility of CRISPR/Cas9 genome engineering on AAV producer cell lines could be further explored to generate potentially safer AAV gene therapy products. This could be achieved by examining the effect of rAAV-FLuc vectors derived from YB1 or alternative knockout cell lines on the induction of an immune response *in vivo*, compared to standard rAAV vectors as control. Additionally, the transduction capacity of these rAAV-FLuc vectors could be quantitatively assessed *in vivo* and also compared to standard rAAV vectors as control. Ultimately, the case for a standardised producer cell line, established by CRISPR/Cas9 genome editing, could be argued in favour for clinical grade rAAV vector production.

7.14 Final conclusions

We utilised CRISPR/Cas9 genome editing of 293T and Sf9 cell lines to improve rAAV vector production. In the process we identified novel Y-Box protein homologues endogenous to *T. ni* and *Spodoptera spp.*, including *S. frugiperda*, from which Sf9 cells are derived. We do note that CRISPR/Cas9 technologies (especially those that waive rights in their intellectual properties for the final CRISPR/Cas9-edited product) designated for insect cells are limited. Such CRISPR/Cas9 technologies are predominantly tailored for *in vitro* use in mammalian cell lines. Therefore, use of IVT sgRNA:rCas9 RNP transfection was preferred over pCRISPR transfections, especially given the CMV promoter restriction in Sf9 cells, and the risk-free genomic integration to give rise to stable Cas9-expressing cell lines. Regardless, we show that knockout of YB1 or SfYB was confidently achieved, and that established cell lines exhibited homozygosity for the observed knockout mutation(s); however, this was dependent on the ability to establish monoclonal cell populations by single cell cloning. We also addressed the advantages of using a broad scope of phenotypic and genotypic analysis tools to characterise knockout cell lines. This included HRM curve analyses, which differentiated wildtype alleles from CRISPR/Cas9-mutated alleles with high sensitivity. We stress the importance of using target-specific antibodies for screening knockout cell lines to avoid potential false-positive or false-negative data interpretations from Western blotting analyses. We were also able to model the disruption of wildtype splicing junctions and that alternative splice sites could be generated as consequences of CRISPR/Cas9-mediated mutations.

YB1 knockout cells were significantly sensitised to chloroquine-induced cell cytotoxicity, but unaffected by alternative transfection or infection methodologies for vector production. A series of YB1 truncation mutants further exemplified the importance of YB1 in protecting 293T cells from chloroquine-induced cell death, and that YB1's CSD was the principle mediator of this resistance. Lastly, further implications in YB1-AAV interactions were indicated, particularly YB1 interactions with AAV2 *ITR* 5'-GGGGTT-3' binding motif and that YB1's CTD colocalised with rAAV2 vector particles and AAV2 Cap protein *in vitro*.

Ultimately, careful consideration in the choice in endogenous gene target for CRISPR/Cas9 genome editing is stressed. Targeting endogenous host-cell proteins may potentially have unexpected consequences for downstream investigations. This ultimately depends on the importance or versatility of the target protein's function *in vitro*. Nonetheless, we present the first example in which CRISPR/Cas9 genome editing can be utilised to regulate cell-intrinsic

factors that may be implicated in rAAV vector production. We also demonstrate the capabilities of harnessing the CRISPR/Cas9 technology for potential in gene therapy product development. With these in mind, CRISPR/Cas9 technology can further revolutionise human gene therapy beyond its current applications and give rise to an entirely new class of therapeutics.

Bibliography

- ABDOLI, A., SOLEIMANJAH, H., FOTOUHI, F., TEIMOORI, A., POUR BEIRANVAND, S. & KIANMEHR, Z. 2013. Human Papillomavirus type 16- L1 VLP production in insect cells. *Iran J Basic Med Sci*, 16, 891-5.
- AGBANDJE-MCKENNA, M. & KLEINSCHMIDT, J. 2011. AAV capsid structure and cell interactions. *Methods Mol Biol*, 807, 47-92.
- AHMAD, M., SRINIVASULA, S. M., WANG, L., LITWACK, G., FERNANDES-ALNEMRI, T. & ALNEMRI, E. S. 1997. Spodoptera frugiperda caspase-1, a novel insect death protease that cleaves the nuclear immunophilin FKBP46, is the target of the baculovirus antiapoptotic protein p35. *J Biol Chem*, 272, 1421-4.
- AHN, M., WITTING, S. R., RUIZ, R., SAXENA, R. & MORRAL, N. 2011. Constitutive expression of short hairpin RNA in vivo triggers buildup of mature hairpin molecules. *Hum Gene Ther*, 22, 1483-97.
- AJIRO, M., JIA, R., WANG, R. H., DENG, C. X. & ZHENG, Z. M. 2015. Adapted resistance to the knockdown effect of shRNA-derived Srsf3 siRNAs in mouse littermates. *Int J Biol Sci*, 11, 1248-56.
- AKINC, A., THOMAS, M., KLIBANOV, A. M. & LANGER, R. 2005. Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis. *J Gene Med*, 7, 657-63.
- ALLEMAND, E., HASTINGS, M. L., MURRAY, M. V., MYERS, M. P. & KRAINER, A. R. 2007. Alternative splicing regulation by interaction of phosphatase PP2Cgamma with nucleic acid-binding protein YB-1. *Nat Struct Mol Biol*, 14, 630-8.
- ALTON, E., ARMSTRONG, D. K., ASHBY, D., BAYFIELD, K. J., BILTON, D., BLOOMFIELD, E. V., BOYD, A. C., BRAND, J., BUCHAN, R., CALCEDO, R., CARVELLI, P., CHAN, M., CHENG, S. H., COLLIE, D. D. S., CUNNINGHAM, S., DAVIDSON, H. E., DAVIES, G., DAVIES, J. C., DAVIES, L. A., DEWAR, M. H., DOHERTY, A., DONOVAN, J., DWYER, N. S., ELGMATI, H. I., FEATHERSTONE, R. F., GAVINO, J., GEA-SORLI, S., GEDDES, D. M., GIBSON, J. S. R., GILL, D. R., GREENING, A. P., GRIESENBACH, U., HANSELL, D. M., HARMAN, K., HIGGINS, T. E., HODGES, S. L., HYDE, S. C., HYNDMAN, L., INNES, J. A., JACOB, J., JONES, N., KEOGH, B. F., LIMBERIS, M. P., LLOYD-EVANS, P., MACLEAN, A. W., MANVELL, M. C., MCCORMICK, D., MCGOVERN, M., MCLACHLAN, G., MENG, C., MONTERO, M. A., MILLIGAN, H., MOYCE, L. J., MURRAY, G. D., NICHOLSON, A. G., OSADOLOR, T., PARRALEITON, J., PORTEOUS, D. J., PRINGLE, I. A., PUNCH, E. K., PYTEL, K. M., QUITTNER, A. L., RIVELLINI, G., SAUNDERS, C. J., SCHEULE, R. K., SHEARD, S., SIMMONDS, N. J., SMITH, K., SMITH, S. N., SOUSSI, N., SOUSSI, S., SPEARING, E. J., STEVENSON, B. J., SUMNER-JONES, S. G., TURKKILA, M., URETA, R. P., WALLER, M. D., WASOWICZ, M. Y., WILSON, J. M., WOLSTENHOLME-HOGG, P. & CONSORTIUM, U. K. C. F. G. T. 2015. Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. *Lancet Respir Med*, 3, 684-691.
- ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-402.
- AMADO, M., ALMEIDA, R., CARNEIRO, F., LEVERY, S. B., HOLMES, E. H., NOMOTO, M., HOLLINGSWORTH, M., HASSAN, H., SCHWIENSTEK, T., NIELSEN, P. A., BENNETT, E. & CLAUSEN, H. 1998. A family of human β 3-galactosyltransferases: Characterization of four members of a UDP-galactose: β -N-acetyl-glucosamine/ β -N-acetyl-galactosamine β -1,3-galactosyltransferase family. *J Biol Chem*, 273, 12770-8.

- ANDERS, C., NIEWOEHNER, O., DUERST, A. & JINEK, M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513, 569-73.
- ANDERSON, W. F. 1984. Prospects for human gene therapy. *Science*, 226, 401-9.
- ANDERSSON, M. G., HAASNOOT, P. C., XU, N., BERENJIAN, S., BERKHOUT, B. & AKUSJARVI, G. 2005. Suppression of RNA interference by adenovirus virus-associated RNA. *J Virol*, 79, 9556-65.
- ANSARI, S. A., SAFAK, M., GALLIA, G. L., SAWAYA, B. E., AMINI, S. & KHALILI, K. 1999. Interaction of YB-1 with human immunodeficiency virus type 1 Tat and TAR RNA modulates viral promoter activity. *J Gen Virol*, 80, 2629-38.
- ASLANIDI, G. V., RIVERS, A. E., ORTIZ, L., SONG, L., LING, C., GOVINDASAMY, L., VAN VLIET, K., TAN, M., AGBANDJE-MCKENNA, M. & SRIVASTAVA, A. 2013. Optimization of the capsid of recombinant adeno-associated virus 2 (AAV2) vectors: the final threshold? *PLoS One*, 8, e59142.
- ASOKAN, A., HAMRA, J. B., GOVINDASAMY, L., AGBANDJE-MCKENNA, M. & SAMULSKI, R. J. 2006. Adeno-associated virus type 2 contains an integrin alpha5beta1 binding domain essential for viral cell entry. *J Virol*, 80, 8961-9.
- ATCHISON, R. W., CASTO, B. C. & HAMMON, W. M. 1965. Adenovirus-associated defective virus particles. *Science*, 149, 754-6.
- AUCOIN, M. G., PERRIER, M. & KAMEN, A. A. 2008. Critical assessment of current adeno-associated viral vector production and quantification methods. *Biotechnol Adv.*, 26, 73-88.
- AVERY, O. T., MACLEOD, C. M. & MCCARTY, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J Exp Med*, 79, 137-58.
- AYRES, M. D., HOWARD, S. C., KUZIO, J., LOPEZ-FERBER, M. & POSSEE, R. D. 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology*, 202, 586-605.
- AYUSO, E., MINGOZZI, F. & BOSCH, F. 2010. Production, purification and characterization of Adeno-associated Vectors. *Current Gene Therapy*, 10, 423-36.
- BADER, A. G. & VOGT, P. K. 2005. Inhibition of protein synthesis by Y box-binding protein 1 blocks oncogenic cell transformation. *Mol Cell Biol*, 25, 2095-106.
- BALAGÚE, C., KALLA, M. & ZHANG, W. W. 1997. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J Virol*, 71, 3299-306.
- BANN, D. V., BEYER, A. R. & PARENT, L. J. 2014. A murine retrovirus co-opts YB-1, a translational regulator and stress granule-associated protein, to facilitate virus assembly. *J Virol*, 88, 4434-50.
- BANTEL-SCHAAL, U., HUB, B. & KARTENBECK, J. 2002. Endocytosis of adeno-associated virus type 5 leads to accumulation of virus particles in the golgi compartment. *Journal of Virology*, 76, 2340-349.
- BANTEL-SCHAAL, U. & ZUR HAUSEN, H. 1984. Characterization of the DNA of a defective human parvovirus isolated from a genital site. *Virology*, 134, 52-63.
- BAR-PELED, L., SCHWEITZER, L. D., ZONCU, R. & SABATINI, D. M. 2012. Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1. *Cell*, 150, 1196-208.
- BARGONETTI, J., REYNISDÓTTIR, I., N FRIEDMAN, P. & PRIVES, C. 1992. Site-specific binding of wild-type p53 to cellular DNA is inhibited by SV40 T antigen and mutant p53. *Genes Dev*, 6, 1886-98.
- BAROTI, T., SCHILLINGER, A., WEGNER, M. & STOLT, C. C. 2016. Sox13 functionally complements the related Sox5 and Sox6 as important developmental modulators in mouse spinal cord oligodendrocytes. *J Neurochem*, 136, 316-28.

- BARRANGOU, R. 2015. Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biol*, 16, 247.
- BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315, 1709-12.
- BARRANGOU, R. & MARRAFFINI, L. A. 2014. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell*, 54, 234-44.
- BARTLETT, D. W. & DAVIS, M. E. 2006. Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. *Nucleic Acids Res*, 34, 322-33.
- BARTLETT, J. S., WILCHER, R. & SAMULSKI, R. J. 2000. Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. *J Virol*, 74, 2777-85.
- BENGTSSON, N. E., HALL, J. K., ODOM, G. L., PHELPS, M. P., ANDRUS, C. R., HAWKINS, R. D., HAUSCHKA, S. D., CHAMBERLAIN, J. R. & CHAMBERLAIN, J. S. 2017. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nature Communications*, 8, 14454.
- BENNETT, J., WELLMAN, J., MARSHALL, K. A., MCCAGUE, S., ASHTARI, M., DISTEFANO-PAPPAS, J., ELCI, O. U., CHUNG, D. C., SUN, J., WRIGHT, J. F., CROSS, D. R., ARAVAND, P., CYCKOWSKI, L. L., BENNICELLI, J. L., MINGOZZI, F., AURICCHIO, A., PIERCE, E. A., RUGGIERO, J., LEROY, B. P., SIMONELLI, F., HIGH, K. A. & MAGUIRE, A. M. 2016. Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. *The Lancet*, 388, 661-72.
- BERNS, K. I., PINKERTON, T. C., THOMAS, G. F. & HOGGAN, M. D. 1975. Detection of adeno-associated virus (AAV)-specific nucleotide sequences in DNA isolated from latently infected Detroit 6 cells. *Virology*, 68, 556-60.
- BERNS, K. I. & ROSE, J. A. 1970. Evidence for a single-stranded adenovirus-associated virus genome: isolation and separation of complementary single strands. *J Virol*, 5, 693-9.
- BERNSTEIN, E., CAUDY, A. A., HAMMOND, S. M. & HANNON, G. J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409, 363-6.
- BERRY, G. E. & ASOKAN, A. 2016. Chemical Modulation of Endocytic Sorting Augments Adeno-associated Viral Transduction. *J Biol Chem*, 291, 939-47.
- BEVINGTON, J. M., NEEDHAM, P. G., VERRILL, K. C., COLLACO, R. F., BASRUR, V. & TREMPE, J. P. 2007. Adeno-associated virus interactions with B23/Nucleophosmin: identification of sub-nucleolar virion regions. *Virology*, 357, 102-13.
- BIBIKOVA, M., GOLIC, M., GOLIC, K. G. & CARROLL, D. 2002. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 161, 1169-75.
- BLANCHETTE, P., KINDSMULLER, K., GROITL, P., DALLAIRE, F., SPEISEDER, T., BRANTON, P. E. & DOBNER, T. 2008. Control of mRNA export by adenovirus E4orf6 and E1B55K proteins during productive infection requires E4orf6 ubiquitin ligase activity. *J Virol*, 82, 2642-51.
- BLANK, C., KUBALL, J., VOELKL, S., WIENDL, H., BECKER, B., WALTER, B., MAJDIC, O., GAJEWSKI, T. F., THEOBALD, M., ANDREESEN, R. & MACKENSEN, A. 2006. Blockade of PD-L1 (B7-H1) augments human tumor-specific T cell responses in vitro. *Int J Cancer*, 119, 317-27.
- BLEKER, S., PAWLITA, M. & KLEINSCHMIDT, J. A. 2006. Impact of capsid conformation and Rep-capsid interactions on adeno-associated virus type 2 genome packaging. *J Virol*, 80, 810-20.
- BLEKER, S., SONNTAG, F. & KLEINSCHMIDT, J. A. 2005. Mutational analysis of narrow pores at the fivefold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase A2 activity. *J Virol*, 79, 2528-40.

- BOCCHETTA, M., ELIASZ, S., DE MARCO, M. A., RUDZINSKI, J., ZHANG, L. & CARBONE, M. 2008. The SV40 large T antigen-p53 complexes bind and activate the insulin-like growth factor-I promoter stimulating cell growth. *Cancer Res*, 68, 1022-9.
- BOLOTIN, A., QUINQUIS, B., SOROKIN, A. & EHRLICH, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151, 2551-61.
- BOUSSIF, O., LEZOUALC'H, F., ZANTA, M. A., DJAVAHARI-MERGNY, M., SCHERMAN, D., DEMENEIX, B. & BEHR, J. P. 1995. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc Natl Acad Sci U S A*, 92, 7297-301.
- BOWLES, D. E., MCPHEE, S. W., LI, C., GRAY, S. J., SAMULSKI, J. J., CAMP, A. S., LI, J., WANG, B., MONAHAN, P. E., RABINOWITZ, J. E., GRIEGER, J. C., GOVINDASAMY, L., AGBANDJE-MCKENNA, M., XIAO, X. & SAMULSKI, R. J. 2012. Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector. *Mol Ther*, 20, 443-55.
- BOYCOTT, K. M., VANSTONE, M. R., BULMAN, D. E. & MACKENZIE, A. E. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, 14, 681-91.
- BRAY, P. G., JANNEH, O., RAYNES, K. J., MUNGTHIN, M., GINSBURG, H. & WARD, S. A. 1999. Cellular uptake of chloroquine is dependent on binding to ferriprotoporphyrin IX and is independent of NHE activity in Plasmodium falciparum. *J Cell Biol*, 145, 363-76.
- BROUNS, S. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J., SNIJDERS, A. P., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321, 960-4.
- BROWN, P. A., BODLES-BRAKHOP, A. M., POPE, M. A. & DRAGHIA-AKLI, R. 2009. Gene therapy by electroporation for the treatment of chronic renal failure in companion animals. *BMC Biotechnol*, 9, 4.
- BRYANT, L. M., CHRISTOPHER, D. M., GILES, A. R., HINDERER, C., RODRIGUEZ, J. L., SMITH, J. B., TRAXLER, E. A., TYCKO, J., WOJNO, A. P. & WILSON, J. M. 2013. Lessons learned from the clinical development and market authorization of Glybera. *Hum Gene Ther Clin Dev*, 24, 55-64.
- BULLER, R. M., JANIK, J. E., SEBRING, E. D. & ROSE, J. A. 1981. Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. *J Virol*, 40, 241-7.
- BUNING, H., HUBER, A., ZHANG, L., MEUMANN, N. & HACKER, U. 2015. Engineering the AAV capsid to optimize vector-host-interactions. *Curr Opin Pharmacol*, 24, 94-104.
- BUNSE, M., BENDLE, G. M., LINNEMANN, C., BIES, L., SCHULZ, S., SCHUMACHER, T. N. & UCKERT, W. 2014. RNAi-mediated TCR knockdown prevents autoimmunity in mice caused by mixed TCR dimers following TCR gene transfer. *Mol Ther*, 22, 1983-91.
- BYLUND, L., KYTÖLÄ, S., LUI, W. O., LARSSON, C. & WEBER, G. 2004. Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches. *Cytogenet Genome Res*, 106, 28-32.
- CALLES, K., SVENSSON, I., LINDSKOG, E. & HAGGSTROM, L. 2006. Effects of conditioned medium factors and passage number on Sf9 cell physiology and productivity. *Biotechnol Prog*, 22, 394-400.
- CAO, M., YOU, H. & HERMONAT, P. L. 2014. The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication. *PLoS One*, 9, e104596.
- CAPECCHI, M. R. 1989. Altering the genome by homologous recombination. *Science*, 244, 1288-92.
- CARINHAS, N., ROBITAILLE, A. M., MOES, S., CARRONDO, M. J., JENOE, P., OLIVEIRA, R. & ALVES, P. M. 2011. Quantitative proteomics of Spodoptera frugiperda cells during growth and baculovirus infection. *PLoS One*, 6, e26444.

- CARPENTIER, A. C., FRISCH, F., LABBE, S. M., GAGNON, R., DE WAL, J., GREENTREE, S., PETRY, H., TWISK, J., BRISSON, D. & GAUDET, D. 2012. Effect of alipogene tiparvovec (AAV1-LPL(S447X)) on postprandial chylomicron metabolism in lipoprotein lipase-deficient patients. *J Clin Endocrinol Metab*, 97, 1635-44.
- CARSTENS, E. B., TJIA, S. T. & DOERFLER, W. 1979. Infection of *Spodoptera frugiperda* cells with *Autographa californica* nuclear polyhedrosis virus I. Synthesis of intracellular proteins after virus infection. *Virology*, 99, 386-98.
- CARTE, J., WANG, R., LI, H., TERNS, R. M. & TERNS, M. P. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev*, 22, 3489-96.
- CAUDRON, N., ARNAL, I., BUHLER, E., JOB, D. & VALIRON, O. 2002. Microtubule nucleation from stable tubulin oligomers. *J Biol Chem*, 277, 50973-9.
- CHANG, L. S. & SHENK, T. 1990. The adenovirus DNA-binding protein stimulates the rate of transcription directed by adenovirus and adeno-associated virus promoters. *J Virol*, 64, 2103-9.
- CHANG, T. H., RAY, F. A., THOMPSON, D. & SCHLEGEL, R. 1997. Disregulation of mitotic checkpoints and regulatory proteins following acute expression of SV40 large T antigen in diploid human cells. *Oncogene*, 14, 2383-93.
- CHANSKY, H. A., HU, M., HICKSTEIN, D. D. & YANG, L. 2001. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer Res*, 61, 3586-90.
- CHARLESWORTH, C. T., DESHPANDE, P. S., DEVER, D. P., DEJENE, B., GOMEZ-OSPINA, N., MANTRI, S., PAVEL-DINU, M., CAMARENA, J., WEINBERG, K. I. & PORTEUS, M. H. 2018. Identification of pre-existing adaptive immunity to Cas9 proteins in humans. *bioRxiv*.
- CHEN, C. & OKAYAMA, H. 1987. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol*, 7, 2747-52.
- CHEN, N. N., CHANG, C. F., GALLIA, G. L., KERR, D., JOHNSON, E. M., KRACHMAROV, C. P., BARR, S. M., FRISQUE, R. J., BOLLAG, B. & KHALILI, K. 1995. Cooperative action of cellular proteins YB-1 and Pur α with the T-antigen of the human polyomavirus JCv determines their interaction with the viral lytic control element. *Proc Natl Acad Sci U S A*, 92, 1087-91.
- CHEN, N. N. & KHALILI, K. 1995. Transcriptional regulation of human JC polyomavirus promoters by cellular proteins YB-1 and Pur α in glial cells. *J Virol*, 69, 5843-8.
- CHEN, Y., ZHU, X., ZHANG, X., LIU, B. & HUANG, L. 2010. Nanoparticles modified with tumor-targeting scFv deliver siRNA and miRNA for cancer therapy. *Mol Ther*, 18, 1650-6.
- CHENDRIMADA, T. P., FINN, K. J., JI, X., BAILLAT, D., GREGORY, R. I., LIEBHABER, S. A., PASQUINELLI, A. E. & SHIEKHATTAR, R. 2007. MicroRNA silencing through RISC recruitment of eIF6. *Nature*, 447, 823-8.
- CHENG, F., GONG, L., ZHAO, D., YANG, H., ZHOU, J., LI, M. & XIANG, H. 2017a. Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *J Genet Genomics*, 44, 541-8.
- CHENG, J., ZEIDAN, R., MISHRA, S., LIU, A., PUN, S., P KULKARNI, R., S JENSEN, G., C BELLOCQ, N. & DAVIS, M. 2006. Structure-function correlation of chloroquine and analogues as transgene expression enhancers in nonviral gene delivery. *J Med Chem*, 49, 6522-31.
- CHENG, T., WU, J., WU, Y., CHILUKURI, R. V., HUANG, L., YAMAMOTO, K., FENG, L., LI, W., CHEN, Z., GUO, H., LIU, J., LI, S., WANG, X., PENG, L., LIU, D., GUO, Y., FU, B., LI, Z., LIU, C., CHEN, Y., TOMAR, A., HILLIOU, F., MONTAGNE, N., JACQUIN-JOLY, E., D'ALENCON, E., SETH, R. K., BHATNAGAR, R. K., JOURAKU, A., SHIOTSUKI, T., KADONO-OKUDA, K., PROMBOON, A., SMAGGHE, G., ARUNKUMAR, K. P., KISHINO, H., GOLDSMITH, M. R., FENG, Q., XIA, Q. & MITA, K. 2017b. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat Ecol Evol*, 1, 1747-56.

- CHERNOV, K. G., MECHULAM, A., POPOVA, N. V., PASTRE, D., NADEZHINA, E. S., SKABKINA, O. V., SHANINA, N. A., VASILIEV, V. D., TARRADE, A., MELKI, J., JOSHI, V., BACONNAIS, S., TOMA, F., OVCHINNIKOV, L. P. & CURMI, P. A. 2008. YB-1 promotes microtubule assembly in vitro through interaction with tubulin and microtubules. *BMC Biochem*, 9, 23.
- CHEW, W. L., TABEBORDBAR, M., CHENG, J. K., MALI, P., WU, E. Y., NG, A. H., ZHU, K., WAGERS, A. J. & CHURCH, G. M. 2016. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods*, 13, 868-74.
- CHOE, H., FARZAN, M., SUN, Y., SULLIVAN, N., ROLLINS, B., PONATH, P. D., WU, L., MACKAY, C. R., LAROSA, G., NEWMAN, W., GERARD, N., GERARD, C. & SODROSKI, J. 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell*, 85, 1135-48.
- CHOULIKA, A., PERRIN, A., DUJON, B. & NICOLAS, J. F. 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol*, 15, 1968-73.
- CHYLINSKI, K., LE RHUN, A. & CHARPENTIER, E. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol*, 10, 726-37.
- COHEN, S., BEHZAD, A. R., CARROLL, J. B. & PANTE, N. 2006. Parvoviral nuclear import: bypassing the host nuclear-transport machinery. *J Gen Virol*, 87, 3209-13.
- COHEN, S., MARR, A. K., GARCIN, P. & PANTE, N. 2011. Nuclear envelope disruption involving host caspases plays a role in the parvovirus replication cycle. *J Virol*, 85, 4863-74.
- COLLACO, R. F., CAO, X. & TREMPER, J. P. 1999. A helper virus-free packaging system for recombinant adeno-associated virus vectors. *Gene*, 238, 397-405.
- COLOMER-LLUCH, M., JOFRE, J. & MUNIESA, M. 2011. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS One*, 6, e17549.
- CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819-23.
- COUTURIER, C., SARKIS, C., SERON, K., BELOUZARD, S., CHEN, P., LENAIN, A., CORSET, L., DAM, J., VAUTHIER, V., DUBART, A., MALLET, J., FROGUEL, P., ROUILLE, Y. & JOCKERS, R. 2007. Silencing of OB-RGRP in mouse hypothalamic arcuate nucleus increases leptin receptor signaling and prevents diet-induced obesity. *Proc Natl Acad Sci U S A*, 104, 19476-81.
- DABERT, P., EHRLICH, S. D. & GRUSS, A. 1992. Chi sequence protects against RecBCD degradation of DNA in vivo. *Proc Natl Acad Sci U S A*, 89, 12073-7.
- DAHL, E., EN-NIA, A., WIESMANN, F., KRINGS, R., DJUDJAJ, S., BREUER, E., FUCHS, T., WILD, P. J., HARTMANN, A., DUNN, S. E. & MERTENS, P. R. 2009. Nuclear detection of Y-box protein-1 (YB-1) closely associates with progesterone receptor negativity and is a strong adverse survival factor in human breast cancer. *BMC Cancer*, 9, 410.
- DANG, Y., JIA, G., CHOI, J., MA, H., ANAYA, E., YE, C., SHANKAR, P. & WU, H. 2015. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol*, 16, 280.
- DAS, S., CHATTOPADHYAY, R., BHAKAT, K. K., BOLDOGH, I., KOHNO, K., PRASAD, R., WILSON, S. H. & HAZRA, T. K. 2007. Stimulation of NEIL2-mediated oxidized base excision repair via YB-1 interaction during oxidative stress. *J Biol Chem*, 282, 28474-84.
- DATSENKO, K. A., POUGACH, K., TIKHONOV, A., WANNER, B. L., SEVERINOV, K. & SEMENOVA, E. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun*, 3, 945.
- DAVYDOVA, E. K., EVDOKIMOVA, V. M., OVCHINNIKOV, L. P. & HERSHEY, J. W. 1997. Overexpression in COS cells of p50, the major core protein associated with mRNA, results in translation inhibition. *Nucleic Acids Res*, 25, 2911-6.

- DAYA, S., CORTEZ, N. & BERNS, K. I. 2009. Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway. *J Virol*, 83, 11655-64.
- DE SOUZA-PINTO, N. C., MASON, P. A., HASHIGUCHI, K., WEISSMAN, L., TIAN, J., GUAY, D., LABEL, M., STEVNSNER, T. V., RASMUSSEN, L. J. & BOHR, V. A. 2009. Novel DNA mismatch-repair activity involving YB-1 in human mitochondria. *DNA Repair (Amst)*, 8, 704-19.
- DECKERT, J., HARTMUTH, K., BOEHRINGER, D., BEHZADNIA, N., WILL, C. L., KASTNER, B., STARK, H., URLAUB, H. & LUHRMANN, R. 2006. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol*, 26, 5528-43.
- DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471, 602-7.
- DELVECCHIO, R., HIGA, L. M., PEZZUTO, P., VALADAO, A. L., GARCEZ, P. P., MONTEIRO, F. L., LOIOLA, E. C., DIAS, A. A., SILVA, F. J., ALIOTA, M. T., CAINE, E. A., OSORIO, J. E., BELLIO, M., O'CONNOR, D. H., REHEN, S., DE AGUIAR, R. S., SAVARINO, A., CAMPANATI, L. & TANURI, A. 2016. Chloroquine, an endocytosis blocking agent, inhibits Zika virus infection in different cell models. *Viruses*, 8, E322.
- DESMET, F. O., HAMROUN, D., LALANDE, M., COLLOD-BEROUD, G., CLAUSTRES, M. & BEROUD, C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*, 37, e67.
- DEY, A., ROBITAILLE, M., REMKE, M., MAIER, C., MALHOTRA, A., GREGORIEFF, A., WRANA, J. L., TAYLOR, M. D., ANGERS, S. & KENNEY, A. M. 2016. YB-1 is elevated in medulloblastoma and drives proliferation in Sonic hedgehog-dependent cerebellar granule neuron progenitor cells and medulloblastoma cells. *Oncogene*, 35, 4256-68.
- DHILLON, J., ASTANEHE, A., LEE, C., FOTOVATI, A., HU, K. & DUNN, S. E. 2010. The expression of activated Y-box binding protein-1 serine 102 mediates trastuzumab resistance in breast cancer cells by increasing CD44+ cells. *Oncogene*, 29, 6294-300.
- DIDIER, D. K., SCHIFFENBAUER, J., WOULFE, S. L., ZACHEIS, M. & SCHWARTZ, B. D. 1988. Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc Natl Acad Sci U S A*, 85, 7322-6.
- DING, W., ZHANG, L. N., YEAMAN, C. & ENGELHARDT, J. F. 2006. rAAV2 traffics through both the late and the recycling endosomes in a dose-dependent fashion. *Mol Ther*, 13, 671-82.
- DOENCH, J. G., FUSI, N., SULLENDER, M., HEGDE, M., VAIMBERG, E. W., DONOVAN, K. F., SMITH, I., TOTHOVA, Z., WILEN, C., ORCHARD, R., VIRGIN, H. W., LISTGARTEN, J. & ROOT, D. E. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, 34, 184-191.
- DOENCH, J. G., HARTENIAN, E., GRAHAM, D. B., TOTHOVA, Z., HEGDE, M., SMITH, I., SULLENDER, M., EBERT, B. L., XAVIER, R. J. & ROOT, D. E. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, 32, 1262-7.
- DOI, N., ZENNO, S., UEDA, R., OHKI-HAMAZAKI, H., UI-TEI, K. & SAIGO, K. 2003. Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr Biol*, 13, 41-6.
- DONG, B., DUAN, X., CHOW, H. Y., CHEN, L., LU, H., WU, W., HAUCK, B., WRIGHT, F., KAPRANOV, P. & XIAO, W. 2014. Proteomics analysis of co-purifying cellular proteins associated with rAAV vectors. *PLoS One*, 9, e86453.

- DONG, Z. Q., HU, N., DONG, F. F., CHEN, T. T., JIANG, Y. M., CHEN, P., LU, C. & PAN, M. H. 2017. Baculovirus LEF-11 Hijack Host ATPase ATAD3A to Promote Virus Multiplication in *Bombyx mori* cells. *Sci Rep*, 7, 46187.
- DONSANTE, A., MILLER, D. G., LI, Y., VOGLER, C., BRUNT, E. M., RUSSELL, D. W. & SANDS, M. S. 2007. AAV vector integration sites in mouse hepatocellular carcinoma. *Science*, 317, 477.
- DONSANTE, A., VOGLER, C., MUZYCZKA, N., CRAWFORD, J. M., BARKER, J., FLOTTE, T., CAMPBELL-THOMPSON, M., DALY, T. & SANDS, M. S. 2001. Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. *Gene Ther*, 8, 1343-6.
- DOUAR, A. M., POULARD, K., STOCKHOLM, D. & DANOS, O. 2001. Intracellular trafficking of adeno-associated virus vectors: routing to the late endosomal compartment and proteasome degradation. *J Virol*, 75, 1824-33.
- DRITTANTI, L., JENNY, C., POULARD, K., SAMBA, A., MANCEAU, P., SORIA, N., VINCENT, N., DANOS, O. & VEGA, M. 2001. Optimised helper virus-free production of high-quality adeno-associated virus vectors. *J Gene Med*, 3, 59-71.
- DUAN, D., LI, Q., KAO, A. W., YUE, Y., PESSIN, J. E. & ENGELHARDT, J. F. 1999. Dynamin is required for recombinant adeno-associated virus type 2 infection. *J Virol*, 73, 10371-6.
- DUAN, Y. Y., WU, J., ZHU, J. L., LIU, S. L., OZAKI, I., STRAYER, D. S. & ZERN, M. A. 2004. Gene therapy for human α 1-antitrypsin deficiency in an animal model using SV40-Derived vectors. *Gastroenterology*, 127, 1222-32.
- DUBIELZIG, R., KING, J. A., WEGER, S., KERN, A. & KLEINSCHMIDT, J. A. 1999. Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. *J Virol*, 73, 8989-98.
- DUBRIDGE, R. B., TANG, P., HSIA, H. C., LEONG, P. M., MILLER, J. H. & CALOS, M. P. 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol*, 7, 379-87.
- EARLEY, L. F., KAWANO, Y., ADACHI, K., SUN, X. X., DAI, M. S. & NAKAI, H. 2015. Identification and characterization of nuclear and nucleolar localization signals in the adeno-associated virus serotype 2 assembly-activating protein. *J Virol*, 89, 3038-48.
- EBINA, H., MISAWA, N., KANEMURA, Y. & KOYANAGI, Y. 2013. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci Rep*, 3, 2510.
- ELBASHIR, S. M., HARBORTH, J., LENDECKEL, W., YALCIN, A., WEBER, K. & TUSCHL, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411, 494-8.
- ELLIS, B. L., HIRSCH, M. L., BARKER, J. C., CONNELLY, J. P., STEININGER, R. J. & PORTEUS, M. H. 2013. A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. *Virology Journal*, 10, 74.
- EVDOKIMOVA, V., RUZANOV, P., IMATAKA, H., RAUGHT, B., SVITKIN, Y., OVCHINNIKOV, L. P. & SONENBERG, N. 2001. The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. *Embo j*, 20, 5491-502.
- EVDOKIMOVA, V. M., KOVRIGINA, E. A., NASHCHEKIN, D. V., DAVYDOVA, E. K., HERSHEY, J. W. & OVCHINNIKOV, L. P. 1998. The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro. *J Biol Chem*, 273, 3574-81.
- EZHEVSKY, S. A., NAGAHARA, H., VOCERO-AKBANI, A. M., GIUS, D. R., WEI, M. C. & DOWDY, S. F. 1997. Hypo-phosphorylation of the retinoblastoma protein (pRb) by cyclin D:Cdk4/6 complexes results in active pRb. *Proc Natl Acad Sci U S A*, 94, 10699-704.
- FELGNER, J. H., KUMAR, R., SRIDHAR, C. N., WHEELER, C. J., TSAI, Y. J., BORDER, R., RAMSEY, P., MARTIN, M. & FELGNER, P. L. 1994. Enhanced gene delivery and mechanism studies with a novel series of cationic lipid formulations. *J Biol Chem*, 269, 2550-61.

- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39, 783-791.
- FENG, L., GUO, M., ZHANG, S., CHU, J., ZHUANG, Y. & ZHANG, S. 2007. Optimization of transfection mediated by calcium phosphate for plasmid rAAV-LacZ (recombinant adeno-associated virus-beta-galactosidase reporter gene) production in suspension-cultured HEK-293 (human embryonic kidney 293) cells. *Biotechnol Appl Biochem*, 46, 127-35.
- FENG, L., GUO, M., ZHANG, S., CHU, J., ZHUANG, Y. & ZHANG, S. 2008. Improvement in the suspension-culture production of recombinant adeno-associated virus-LacZ in HEK-293 cells using polyethyleneimine-DNA complexes in combination with hypothermic treatment. *Biotechnol Appl Biochem*, 50, 121-32.
- FIESER, T. M., TAINER, J. A., GEYSEN, H. M., HOUGHTEN, R. A. & LERNER, R. A. 1987. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. *Proc Natl Acad Sci U S A*, 84, 8568-72.
- FLOREA, B. I., MEANEY, C., JUNGINGER, H. E. & BORCHARD, G. 2002. Transfection efficiency and toxicity of polyethylenimine in differentiated Calu-3 and nondifferentiated COS-1 cell cultures. *AAPS PharmSci*, 4, E12.
- FOLGER, K. R., WONG, E. A., WAHL, G. & CAPECCHI, M. R. 1982. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol Cell Biol*, 2, 1372-87.
- FONFARA, I., RICHTER, H., BRATOVIC, M., LE RHUN, A. & CHARPENTIER, E. 2016. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, 532, 517-21.
- FRATTINI, A., FABBRI, M., VALLI, R., DE PAOLI, E., MONTALBANO, G., GRIBALDO, L., PASQUALI, F. & MASERATI, E. 2015. High variability of genomic instability and gene expression profiling in different HeLa clones. *Sci Rep*, 5, 15377.
- FRENCH, R. A., SAMELSON-JONES, B. J., NIEMEYER, G. P., LOTHROP, C. D., JR., MERRICKS, E. P., NICHOLS, T. C. & ARRUDA, V. R. 2018. Complete correction of hemophilia B phenotype by FIX-Padua skeletal muscle gene therapy in an inhibitor-prone dog model. *Blood Adv*, 2, 505-8.
- FRIEBOES, H. B., HUANG, J. S., YIN, W. C. & MCNALLY, L. R. 2014. Chloroquine-mediated cell death in metastatic pancreatic adenocarcinoma through inhibition of autophagy. *Jop*, 15, 189-97.
- FRISQUE, R. J., BREAM, G. L. & CANNELLA, M. T. 1984. Human polyomavirus JC virus genome. *J Virol*, 51, 458-69.
- FRYE, B. C., HALFTER, S., DJUDJAJ, S., MUEHLENBERG, P., WEBER, S., RAFFETSEDER, U., EN-NIA, A., KNOTT, H., BARON, J. M., DOOLEY, S., BERNHAGEN, J. & MERTENS, P. R. 2009. Y-box protein-1 is actively secreted through a non-classical pathway and acts as an extracellular mitogen. *EMBO Rep*, 10, 783-9.
- FU, Y., SANDER, J. D., REYON, D., CASCIO, V. M. & JOUNG, J. K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*, 32, 279-84.
- FUJII, M., KAWAI, K., EGAMI, Y. & ARAKI, N. 2013. Dissecting the roles of Rac1 activation and deactivation in macropinocytosis using microscopic photo-manipulation. *Sci Rep*, 3, 2385.
- FUJITA, T., ITO, K., IZUMI, H., KIMURA, M., SANO, M., NAKAGOMI, H., MAENO, K., HAMA, Y., SHINGU, K., TSUCHIYA, S., KOHNO, K. & FUJIMORI, M. 2005. Increased nuclear localization of transcription factor Y-box binding protein 1 accompanied by up-regulation of P-glycoprotein in breast cancer pretreated with paclitaxel. *Clin Cancer Res*, 11, 8837-44.
- FUJIWARA-OKADA, Y., MATSUMOTO, Y., FUKUSHI, J., SETSU, N., MATSUURA, S., KAMURA, S., FUJIWARA, T., IIDA, K., HATANNO, M., NABESHIMA, A., YAMADA, H., ONO, M., ODA, Y.

- & IWAMOTO, Y. 2013. Y-box binding protein-1 regulates cell proliferation and is associated with clinical outcomes of osteosarcoma. *Br J Cancer*, 108, 836-47.
- GANEM, D., NUSSBAUM, A. L., DAVOLI, D. & FAREED, G. C. 1976. Propagation of a segment of bacteriophage lambda-DNA in monkey cells after covalent linkage to a defective simian virus 40 genome. *Cell*, 7, 349-59.
- GAO, G., VANDENBERGHE, L. H., ALVIRA, M. R., LU, Y., CALCEDO, R., ZHOU, X. & WILSON, J. M. 2004. Clades of adeno-associated viruses are widely disseminated in human tissues. *J Virol*, 78, 6381-8.
- GAO, G. P., ALVIRA, M. R., WANG, L., CALCEDO, R., JOHNSTON, J. & WILSON, J. M. 2002a. Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc Natl Acad Sci U S A*, 99, 11854-9.
- GAO, G. P., LU, F., SANMIGUEL, J. C., TRAN, P. T., ABBAS, Z., LYND, K. S., MARSH, J., SPINNER, N. B. & WILSON, J. M. 2002b. Rep/Cap gene amplification and high-yield production of AAV in an A549 cell line expressing Rep/Cap. *Mol Ther*, 5, 644-9.
- GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*, 109, E2579-86.
- GAUDREULT, I., GUAY, D. & LEBEL, M. 2004. YB-1 promotes strand separation in vitro of duplex DNA containing either mispaired bases or cisplatin modifications, exhibits endonucleolytic activities and binds several DNA repair proteins. *Nucleic Acids Res*, 32, 316-27.
- GE, J., AN, Q., GAO, D., LIU, Y. & PING, W. 2016. Construction of recombinant baculoviruses expressing hemagglutinin of H5N1 avian influenza and research on the immunogenicity. *Sci Rep*, 6, 24290.
- GENG, Y., KOHLI, L., KLOCKE, B. J. & ROTH, K. A. 2010. Chloroquine-induced autophagic vacuole accumulation and cell death in glioma cells is p53 independent. *Neuro Oncol*, 12, 473-81.
- GENOMES PROJECT, C., ABECASIS, G. R., AUTON, A., BROOKS, L. D., DEPRISTO, M. A., DURBIN, R. M., HANDSAKER, R. E., KANG, H. M., MARTH, G. T. & MCVEAN, G. A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56-65.
- GENOMES PROJECT, C., AUTON, A., BROOKS, L. D., DURBIN, R. M., GARRISON, E. P., KANG, H. M., KORBEL, J. O., MARCHINI, J. L., MCCARTHY, S., MCVEAN, G. A. & ABECASIS, G. R. 2015. A global reference for human genetic variation. *Nature*, 526, 68-74.
- GIERMAN, H. J., INDEMANS, M. H., KOSTER, J., GOETZE, S., SEPPEN, J., GEERTS, D., VAN DRIEL, R. & VERSTEEG, R. 2007. Domain-wide regulation of gene expression in the human genome. *Genome Res*, 17, 1286-95.
- GIL-FARINA, I., FRONZA, R., KAEPPPEL, C., LOPEZ-FRANCO, E., FERREIRA, V., D'AVOLA, D., BENITO, A., PRIETO, J., PETRY, H., GONZALEZ-ASEGUINOLAZA, G. & SCHMIDT, M. 2016. Recombinant AAV Integration Is Not Associated With Hepatic Genotoxicity in Nonhuman Primates and Patients. *Mol Ther*, 24, 1100-1105.
- GILBERT, L. A., LARSON, M. H., MORSUT, L., LIU, Z., BRAR, G. A., TORRES, S. E., STERNGINOSSAR, N., BRANDMAN, O., WHITEHEAD, E. H., DOUDNA, J. A., LIM, W. A., WEISSMAN, J. S. & QI, L. S. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154, 442-51.
- GILLIAN-DANIEL, D. L., GRAY, N. K., ÅSTRÖM, J., BARKOFF, A. & WICKENS, M. 1998. Modifications of the 5' cap of mRNAs during *Xenopus* oocyte maturation: independence from changes in poly(A) length and impact on translation. *Molecular and Cellular Biology*, 18, 6152-63.
- GIMENEZ-BONAFE, P., FEDORUK, M. N., WHITMORE, T. G., AKBARI, M., RALPH, J. L., ETTINGER, S., GLEAVE, M. E. & NELSON, C. C. 2004. YB-1 is upregulated during prostate cancer tumor progression and increases P-glycoprotein activity. *Prostate*, 59, 337-49.

- GINSBURG, H. & DEMEL, R. A. 1983. The effect of ferriprotoporphyrin IX and chloroquine on phospholipid monolayers and the possible implications to antimalarial activity. *Biochim Biophys Acta*, 732, 316-9.
- GIROD, A., WOBUS, C. E., ZADORI, Z., RIED, M., LEIKE, K., TIJSSEN, P., KLEINSCHMIDT, J. A. & HALLEK, M. 2002. The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J Gen Virol*, 83, 973-8.
- GLUZMAN, Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell*, 23, 175-82.
- GOFF, S. P. & BERG, P. 1976. Construction of hybrid viruses containing SV40 and lambda phage DNA segments and their propagation in cultured monkey cells. *Cell*, 9, 695-705.
- GOLDSMITH, M. E., MADDEN, M. J., MORROW, C. S. & COWAN, K. H. 1993. A Y-box consensus sequence is required for basal expression of the human multidrug resistance (mdr1) gene. *J Biol Chem*, 268, 5856-60.
- GOLDSTEIN, H., PETTOELLO-MANTOVANI, M., ANDERSON, C. M., CORDELIER, P., POMERANTZ, R. J. & STRAYER, D. S. 2002. Gene therapy using a simian virus 40-derived vector inhibits the development of in vivo human immunodeficiency virus type 1 infection of severe combined immunodeficiency mice implanted with human fetal thymic and liver tissue. *J Infect Dis*, 185, 1425-30.
- GOMEZ-SEBASTIAN, S., LOPEZ-VIDAL, J. & ESCRIBANO, J. M. 2014. Significant productivity improvement of the baculovirus expression vector system by engineering a novel expression cassette. *PLoS One*, 9, e96562.
- GONCALVES, M. A. 2005. Adeno-associated virus: from defective virus to effective vector. *Virology*, 2, 43.
- GONG, B., SHIN, M., SUN, J., JUNG, C. H., BOLT, E. L., VAN DER OOST, J. & KIM, J. S. 2014. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc Natl Acad Sci U S A*, 111, 16359-64.
- GOREN, M. G., DORON, S., GLOBUS, R., AMITAI, G., SOREK, R. & QIMRON, U. 2016. Repeat size determination by two molecular rulers in the type I-E CRISPR array. *Cell Rep*, 16, 2811-8.
- GOUIN, A., BRETAUDEAU, A., NAM, K., GIMENEZ, S., AURY, J. M., DUVIC, B., HILLIOU, F., DURAND, N., MONTAGNE, N., DARBOUX, I., KUWAR, S., CHERTEMPS, T., SIAUSSAT, D., BRETSCHNEIDER, A., MONE, Y., AHN, S. J., HANNIGER, S., GRENET, A. G., NEUNEMANN, D., MAUMUS, F., LUYTEN, I., LABADIE, K., XU, W., KOUTROUMPA, F., ESCOUBAS, J. M., LLOPIS, A., MAIBECHE-COISNE, M., SALASC, F., TOMAR, A., ANDERSON, A. R., KHAN, S. A., DUMAS, P., ORSUCCI, M., GUY, J., BELSER, C., ALBERTI, A., NOEL, B., COULOUX, A., MERCIER, J., NIDELET, S., DUBOIS, E., LIU, N. Y., BOULOGNE, I., MIRABEAU, O., LE GOFF, G., GORDON, K., OAKESHOTT, J., CONSOLI, F. L., VOLKOFF, A. N., FESCEMYER, H. W., MARDEN, J. H., LUTHE, D. S., HERRERO, S., HECKEL, D. G., WINCKER, P., KERGOAT, G. J., AMSELEM, J., QUESNEVILLE, H., GROOT, A. T., JACQUIN-JOLY, E., NEGRE, N., LEMAITRE, C., LEGEAI, F., D'ALENCON, E. & FOURNIER, P. 2017. Two genomes of highly polyphagous lepidopteran pests (Spodoptera frugiperda, Noctuidae) with different host-plant ranges. *Sci Rep*, 7, 11816.
- GRABER, S. G., FIGLER, R. A. & GARRISON, J. C. 1992. Expression and purification of functional G protein alpha subunits using a baculovirus expression system. *J Biol Chem*, 267, 1271-8.
- GRAHAM, F. L., SMILEY, J., RUSSELL, W. C. & NAIRN, R. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*, 36, 59-74.
- GRIEGER, J. C., JOHNSON, J. S., GURDA-WHITAKER, B., AGBANDJE-MCKENNA, M. & SAMULSKI, R. J. 2007. Surface-exposed adeno-associated virus Vp1-NLS capsid fusion protein rescues infectivity of noninfectious wild-type Vp2/Vp3 and Vp3-only capsids but not that of fivefold pore mutant virions. *J Virol*, 81, 7833-43.

- GRIEGER, J. C., SNOWDY, S. & SAMULSKI, R. J. 2006. Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly. *J Virol*, 80, 5199-210.
- GRIEGER, J. C., SOLTYS, S. M. & SAMULSKI, R. J. 2016. Production of recombinant adeno-associated virusvectors using suspension HEK293 cells and continuous harvest of vector from the culture media for GMP FIX and FLT1 clinical vector. *Mol Ther*, 24, 287-97.
- GRIFFITH, F. 1928. The significance of Pneumococcal types. *J Hyg*, 27, 113-59.
- GRIMSEY, N. J., CORONEL, L. J., CORDOVA, I. C. & TREJO, J. 2016. Recycling and endosomal sorting of protease-activated receptor-1 is distinctly regulated by Rab11A and Rab11B proteins. *J Biol Chem*, 291, 2223-36.
- GRISSA, I., VERGNAUD, G. & POURCEL, C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, 8, 172.
- GUAN, Y., MA, Y., LI, Q., SUN, Z., MA, L., WU, L., WANG, L., ZENG, L., SHAO, Y., CHEN, Y., MA, N., LU, W., HU, K., HAN, H., YU, Y., HUANG, Y., LIU, M. & LI, D. 2016. CRISPR/Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse. *EMBO Mol Med*, 8, 477-88.
- GUILINGER, J. P., THOMPSON, D. B. & LIU, D. R. 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol*, 32, 577-82.
- GUO, T., WANG, S., GUO, X. & LU, C. 2005. Productive infection of Autographa californica nucleopolyhedrovirus in silkworm Bombyx mori strain Haoyue due to the absence of a host antiviral factor. *Virology*, 341, 231-237.
- GURYANOV, S. G., FILIMONOV, V. V., TIMCHENKO, A. A., MELNIK, B. S., KIHARA, H., KUTYSHENKO, V. P., OVCHINNIKOV, L. P. & SEMISOTNOV, G. V. 2013. The major mRNP protein YB-1: structural and association properties in solution. *Biochim Biophys Acta*, 1834, 559-67.
- GURYANOV, S. G., SELIVANOVA, O. M., NIKULIN, A. D., ENIN, G. A., MELNIK, B. S., KRETOV, D. A., SERDYUK, I. N. & OVCHINNIKOV, L. P. 2012. Formation of amyloid-like fibrils by Y-box binding protein 1 (YB-1) is mediated by its cold shock domain and modulated by disordered terminal domains. *PLoS One*, 7, e36969.
- HAFT, D. H., SELENGUT, J., MONGODIN, E. F. & NELSON, K. E. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*, 1, e60.
- HALL, A., LARSEN, A. K., PARHAMIFAR, L., MEYLE, K. D., WU, L. P. & MOGHIMI, S. M. 2013. High resolution respirometry analysis of polyethylenimine-mediated mitochondrial energy crisis and cellular stress: Mitochondrial proton leak and inhibition of the electron transport system. *Biochim Biophys Acta*, 1827, 1213-25.
- HALPERN, D., CHIAPELLO, H., SCHBATH, S., ROBIN, S., HENNEQUET-ANTIER, C., GRUSS, A. & EL KAROUI, M. 2007. Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genet*, 3, 1614-21.
- HALVORSEN, O. & SKREDE, S. 1982. Regulation of the biosynthesis of CoA at the level of pantothenate kinase. *Eur J Biochem*, 124, 211-5.
- HAMMOND, S. M., BERNSTEIN, E., BEACH, D. & HANNON, G. J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404, 293-6.
- HANSSEN, L., ALIDOUSTY, C., DJUDJAJ, S., FRYE, B. C., RAUEN, T., BOOR, P., MERTENS, P. R., VAN ROEYEN, C. R., TACKE, F., HEYMANN, F., TITTEL, A. P., KOCH, A., FLOEGE, J., OSTENDORF, T. & RAFFETSEDER, U. 2013. YB-1 is an early and central mediator of bacterial and sterile inflammation in vivo. *J Immunol*, 191, 2604-13.

- HARBISON, C. E., LYI, S. M., WEICHERT, W. S. & PARRISH, C. R. 2009. Early steps in cell infection by parvoviruses: host-specific differences in cell receptor binding but similar endosomal trafficking. *J Virol*, 83, 10504-14.
- HARGROVE, P. W., KEPES, S., HANAHA, H., OBENAUER, J. C., PEI, D., CHENG, C., GRAY, J. T., NEALE, G. & PERSONS, D. A. 2008. Globin lentiviral vector insertions can perturb the expression of endogenous genes in β -thalassemic hematopoietic cells. *Molecular Therapy*, 16, 525-33.
- HARTMUTH, K., URLAUB, H., VORNLOCHER, H. P., WILL, C. L., GENTZEL, M., WILM, M. & LUHRMANN, R. 2002. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A*, 99, 16719-24.
- HASAN, M. T., SUBBAROYAN, R. & CHANG, T. Y. 1991. High-efficiency stable gene transfection using chloroquine-treated Chinese hamster ovary cells. *Somat Cell Mol Genet*, 17, 513-7.
- HASEGAWA, S. L., DOETSCH, P. W., HAMILTON, K. K., MARTIN, A. M., OKENQUIST, S. A., LENZ, J. & BOSS, J. M. 1991. DNA binding properties of YB-1 and dbpA: binding to doublestranded, single-stranded, and abasic site containing DNAs. *Nucleic Acids Research*, 19, 4915-20.
- HASHIMOTO, Y., ZHANG, S. & BLISSARD, G. W. 2010. Ao38, a new cell line from eggs of the black witch moth, *Ascalapha odorata* (Lepidoptera: Noctuidae), is permissive for AcMNPV infection and produces high levels of recombinant proteins. *BMC Biotechnol*, 10, 50.
- HASHIMOTO, Y., ZHANG, S., ZHANG, S., CHEN, Y. R. & BLISSARD, G. W. 2012. Correction: BTI-Tnao38, a new cell line derived from *Trichoplusia ni*, is permissive for AcMNPV infection and produces high levels of recombinant proteins. *BMC Biotechnol*, 12, 12.
- HE, F., HO, Y., YU, L. & KWANG, J. 2008. WSSV ie1 promoter is more efficient than CMV promoter to express H5 hemagglutinin from influenza virus in baculovirus as a chicken vaccine. *BMC Microbiol*, 8, 238.
- HEINZEL, S. S., KRYSAN, P. J., CALOS, M. P. & DUBRIDGE, R. B. 1988. Use of simian virus 40 replication to amplify Epstein-Barr virus shuttle vectors in human cells. *J Virol*, 62, 3738-46.
- HELER, R., SAMAI, P., MODELL, J. W., WEINER, C., GOLDBERG, G. W., BIKARD, D. & MARRAFFINI, L. A. 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature*, 519, 199-202.
- HENCKAERTS, E., DUTHEIL, N., ZELTNER, N., KATTMAN, S., KOHLBRENNER, E., WARD, P., CLEMENT, N., REBOLLO, P., KENNEDY, M., KELLER, G. M. & LINDEN, R. M. 2009. Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proc Natl Acad Sci U S A*, 106, 7571-6.
- HENSSEN, A. G., HENAFF, E., JIANG, E., EISENBERG, A. R., CARSON, J. R., VILLASANTE, C. M., RAY, M., STILL, E., BURNS, M., GANDARA, J., FESCHOTTE, C., MASON, C. E. & KENTISIS, A. 2015. Genomic DNA transposition induced by human PGBD5. *Elife*, 4.
- HERMANNSTADTER, A., ZIEGLER, C., KUHL, M., DEPERT, W. & TOLSTONOG, G. V. 2009. Wild-type p53 enhances efficiency of simian virus 40 large-T-antigen-induced cellular transformation. *J Virol*, 83, 10106-18.
- HEUSINGER, E., KLUGE, S. F., KIRCHHOFF, F. & SAUTER, D. 2015. Early vertebrate evolution of the host restriction factor tetherin. *J Virol*, 89, 12154-65.
- HICKMAN, A. B., RONNING, D. R., PEREZ, Z. N., KOTIN, R. M. & DYDA, F. 2004. The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol Cell*, 13, 403-14.
- HIGASHI, K., INAGAKI, Y., SUZUKI, N., MITSUI, S., MAUVIEL, A., KANEKO, H. & NAKATSUKA, I. 2003. Y-box-binding protein YB-1 mediates transcriptional repression of human alpha 2(I) collagen gene expression by interferon-gamma. *J Biol Chem*, 278, 5156-62.

- HILDINGER, M., BALDI, L., STETTLER, M. & WURM, F. M. 2007. High-titer, serum-free production of adeno-associated virus vectors by polyethyleneimine-mediated plasmid transfection in mammalian suspension cells. *Biotechnol Lett*, 29, 1713-21.
- HOFFMANN, M., WU, Y. J., GERBER, M., BERGER-RENTSCH, M., HEIMRICH, B., SCHWEMMLE, M. & ZIMMER, G. 2010. Fusion-active glycoprotein G mediates the cytotoxicity of vesicular stomatitis virus M mutants lacking host shut-off activity. *J Gen Virol*, 91, 2782-93.
- HOGGAN, M. D., BLACKLOW, N. R. & ROWE, W. P. 1966. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc Natl Acad Sci U S A*, 55, 1467-74.
- HOLM, P. S., BERGMANN, S., JURCHOTT, K., LAGE, H., BRAND, K., LADHOFF, A., MANTWILL, K., CUIEL, D. T., DOBBELSTEIN, M., DIETEL, M., GANSBACHER, B. & ROYER, H. D. 2002. YB-1 relocates to the nucleus in adenovirus-infected cells and facilitates viral replication by inducing E2 gene expression through the E2 late promoter. *J Biol Chem*, 277, 10427-34.
- HOLSCHER, C., SONNTAG, F., HENRICH, K., CHEN, Q., BENEKE, J., MATULA, P., ROHR, K., KADERALI, L., BEIL, N., ERFLE, H., KLEINSCHMIDT, J. A. & MULLER, M. 2015. The SUMOylation pathway restricts gene transduction by adeno-associated viruses. *PLoS Pathog*, 11, e1005281.
- HOMER, C., KNIGHT, D. A., HANANEIA, L., SHEARD, P., RISK, J., LASHAM, A., ROYDS, J. A. & BRAITHWAITE, A. W. 2005. Y-box factor YB1 controls p53 apoptotic function. *Oncogene*, 24, 8314-25.
- HONG, G., WARD, P. & BERNS, K. I. 1994. Intermediates of adeno-associated virus DNA replication in vitro. *J Virol*, 68, 2011-5.
- HOPKINS, R. & ESPOSITO, D. 2009. A rapid method for titrating baculovirus stocks using the Sf-9 Easy Titer cell line. *Biotechniques*, 47, 785-8.
- HOWARD, D. B., POWERS, K., WANG, Y. & HARVEY, B. K. 2008. Tropism and toxicity of adeno-associated viral vector serotypes 1, 2, 5, 6, 7, 8, and 9 in rat neurons and glia in vitro. *Virology*, 372, 24-34.
- HU, W., KAMINSKI, R., YANG, F., ZHANG, Y., COSENTINO, L., LI, F., LUO, B., ALVAREZ-CARBONELL, D., GARCIA-MESA, Y., KARN, J., MO, X. & KHALILI, K. 2014. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc Natl Acad Sci U S A*, 111, 11461-6.
- HUAI, C., JIA, C., SUN, R., XU, P., MIN, T., WANG, Q., ZHENG, C., CHEN, H. & LU, D. 2017. CRISPR/Cas9-mediated somatic and germline gene correction to restore hemostasis in hemophilia B mice. *Hum Genet*, 136, 875-83.
- HUANG, J., CHEN, M., WHITLEY, M. J., KUO, H. C., XU, E. S., WALENS, A., MOWERY, Y. M., VAN MATER, D., EWARD, W. C., CARDONA, D. M., LUO, L., MA, Y., LOPEZ, O. M., NELSON, C. E., ROBINSON-HAMM, J. N., REDDY, A., DAVE, S. S., GERSBACH, C. A., DODD, R. D. & KIRSCH, D. G. 2017. Generation and comparison of CRISPR-Cas9 and Cre-mediated genetically engineered mouse models of sarcoma. *Nat Commun*, 8, 15999.
- HUANG, X., HARTLEY, A. V., YIN, Y., HERSKOWITZ, J. H., LAH, J. J. & RESSLER, K. J. 2013. AAV2 production with optimized N/P ratio and PEI-mediated transfection results in low toxicity and high titer for in vitro and in vivo applications. *J Virol Methods*, 193, 270-7.
- HUIJUN, T., XIANGQING, W. & SENYANG, L. 2014. Construction of a recombinant AAV vector encoding human alpha-synuclein gene with myelin basic protein promoter. *Life Science Journal*, 11, 214-9.
- HUTTER, G., NOWAK, D., MOSSNER, M., GANEPOLA, S., MUSSIG, A., ALLERS, K., SCHNEIDER, T., HOFMANN, J., KUCHERER, C., BLAU, O., BLAU, I. W., HOFMANN, W. K. & THIEL, E. 2009. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med*, 360, 692-8.

- INAGAKI, K., PIAO, C., KOTCHEY, N. M., WU, X. & NAKAI, H. 2008. Frequency and spectrum of genomic integration of recombinant adeno-associated virus serotype 8 vector in neonatal mouse liver. *J Virol*, 82, 9513-24.
- INUZUKA, M., HAYAKAWA, M. & INGI, T. 2005. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *J Biol Chem*, 280, 35776-83.
- ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 169, 5429-33.
- ISHIZU, T., HIGO, S., MASUMURA, Y., KOHAMA, Y., SHIBA, M., HIGO, T., SHIBAMOTO, M., NAKAGAWA, A., MORIMOTO, S., TAKASHIMA, S., HIKOSO, S. & SAKATA, Y. 2017. Targeted genome replacement via homology-directed repair in non-dividing cardiomyocytes. *Sci Rep*, 7, 9363.
- IVANOV, P., EMARA, M. M., VILLEN, J., GYGI, S. P. & ANDERSON, P. 2011. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell*, 43, 613-23.
- IVANOV, P., O'DAY, E., EMARA, M. M., WAGNER, G., LIEBERMAN, J. & ANDERSON, P. 2014. G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proc Natl Acad Sci U S A*, 111, 18201-6.
- IZUMI, H., IMAMURA, T., NAGATANI, G., ISE, T., MURAKAMI, T., URAMOTO, H., TORIGOE, T., ISHIGUCHI, H., YOSHIDA, Y., NOMOTO, M., OKAMOTO, T., UCHIUMI, T., KUWANO, M., FUNA, K. & KOHNO, K. 2001. Y box-binding protein-1 binds preferentially to single-stranded nucleic acids and exhibits 3'→5' exonuclease activity. *Nucleic Acids Res*, 29, 1200-7.
- JACKSON, R. N., GOLDEN, S. M., VAN ERP, P. B., CARTER, J., WESTRA, E. R., BROUNS, S. J., VAN DER OOST, J., TERWILLIGER, T. C., READ, R. J. & WIEDENHEFT, B. 2014. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science*, 345, 1473-9.
- JANELIDZE, S., NORDSTROM, U., KUGLER, S. & BRUNDIN, P. 2014. Pre-existing immunity to adeno-associated virus (AAV)2 limits transgene expression following intracerebral AAV2-based gene delivery in a 6-hydroxydopamine model of Parkinson's disease. *J Gene Med*, 16, 300-8.
- JANOVITZ, T., KLEIN, I. A., OLIVEIRA, T., MUKHERJEE, P., NUSSENZWEIG, M. C., SADELAIN, M. & FALCK-PEDERSEN, E. 2013. High-throughput sequencing reveals principles of adeno-associated virus serotype 2 integration. *J Virol*, 87, 8559-68.
- JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOOLS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43, 1565-75.
- JARMAN-SMITH, R. F., ARMSTRONG, S. J., MANNIX, C. J. & AL-RUBEAI, M. 2002. Chromosome instability in *Spodoptera frugiperda* Sf-9 cell line. *Biotechnol Prog*, 18, 623-8.
- JAY, F. T., DE LA MAZA, L. M. & CARTER, B. J. 1979. Parvovirus RNA transcripts containing sequences not present in mature mRNA: a method for isolation of putative mRNA precursor sequences. *Proceedings of the National Academy of Sciences*, 76, 625-9.
- JIANG, F., ZHOU, K., MA, L., GRESSEL, S. & DOUDNA, J. A. 2015. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, 348, 1477-81.
- JINEK, M., CHYLINSKI, K., INES, F., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-21.
- JINEK, M., JIANG, F., TAYLOR, D. W., STERNBERG, S. H., KAYA, E., MA, E., ANDERS, C., HAUER, M., ZHOU, K., LIN, S., KAPLAN, M., IAVARONE, A. T., CHARPENTIER, E., NOGALES, E. & DOUDNA, J. A. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343, 1247997.

- JOHNSON, J. S., LI, C., DIPRIMIO, N., WEINBERG, M. S., MCCOWN, T. J. & SAMULSKI, R. J. 2010. Mutagenesis of adeno-associated virus type 2 capsid protein VP1 uncovers new roles for basic amino acids in trafficking and cell-specific transduction. *J Virol*, 84, 8888-902.
- JOHNSON, J. S. & SAMULSKI, R. J. 2009. Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. *J Virol*, 83, 2632-44.
- JONES, D. T. & COZZETTO, D. 2015. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*, 31, 857-63.
- JORE, M. M., LUNDGREN, M., VAN DUIJN, E., BULTEMA, J. B., WESTRA, E. R., WAGHMARE, S. P., WIEDENHEFT, B., PUL, U., WURM, R., WAGNER, R., BEIJER, M. R., BARENDREGT, A., ZHOU, K., SNIJDERS, A. P., DICKMAN, M. J., DOUDNA, J. A., BOEKEMA, E. J., HECK, A. J., VAN DER OOST, J. & BROUNS, S. J. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*, 18, 529-36.
- JOSHI, L., DAVIS, T. R., MATTU, T. S., RUDD, P. M., DWEK, R. A., SHULER, M. L. & WOOD, H. A. 2000. Influence of baculovirus-host cell interactions on complex N-linked glycosylation of a recombinant human protein. *Biotechnol Prog*, 16, 650-6.
- JURCHOTT, K., BERGMANN, S., STEIN, U., WALTHER, W., JANZ, M., MANNI, I., PIAGGIO, G., FIETZE, E., DIETEL, M. & ROYER, H. D. 2003. YB-1 as a cell cycle-regulated transcription factor facilitating cyclin A and cyclin B1 gene expression. *J Biol Chem*, 278, 27988-96.
- KAFIL, V. & OMIDI, Y. 2011. Cytotoxic impacts of linear and branched polyethylenimine nanostructures in a431 cells. *Bioimpacts*, 1, 23-30.
- KAKUMANI, P. K., MALHOTRA, P., MUKHERJEE, S. K. & BHATNAGAR, R. K. 2014. A draft genome assembly of the army worm, *Spodoptera frugiperda*. *Genomics*, 104, 134-43.
- KAMINSKI, R., BELLA, R., YIN, C., OTTE, J., FERRANTE, P., GENDELMAN, H. E., LI, H., BOOZE, R., GORDON, J., HU, W. & KHALILI, K. 2016. Excision of HIV-1 DNA by gene editing: a proof-of-concept in vivo study. *Gene Ther*, 23, 696.
- KANG, H., MINDER, P., PARK, M. A., MESQUITTA, W. T., TORBETT, B. E. & SLUKVIN, I. I. 2015. CCR5 disruption in induced pluripotent stem cells using CRISPR/Cas9 provides selective resistance of immune cells to CCR5-tropic HIV-1 virus. *Mol Ther Nucleic Acids*, 4, e268.
- KANG, W., WANG, L., HARRELL, H., LIU, J., THOMAS, D. L., MAYFIELD, T. L., SCOTTI, M. M., YE, G. J., VERES, G. & KNOP, D. R. 2009. An efficient rHSV-based complementation system for the production of multiple rAAV vector serotypes. *Gene Ther*, 16, 229-39.
- KAPAHNKE, M., BANNING, A. & TIKKANEN, R. 2016. Random splicing of several exons caused by a single base change in the target exon of CRISPR/Cas9 mediated gene knockout. *Cells*, 5.
- KARIKO, K., KUO, A., BARNATHAN, E. S. & LANGER, D. J. 1998. Phosphate-enhanced transfection of cationic lipid-complexed mRNA and plasmid DNA. *Biochim Biophys Acta*, 1369, 320-34.
- KARLSON, D. & IMAI, R. 2003. Conservation of the cold shock domain protein family in plants. *Plant Physiol*, 131, 12-5.
- KARVELIS, T., GASIUNAS, G., MIKSYS, A., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2013. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol*, 10, 841-51.
- KARVELIS, T., GASIUNAS, G., YOUNG, J., BIGELYTE, G., SILANSKAS, A., CIGAN, M. & SIKSNYS, V. 2015. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol*, 16, 253.
- KASHIHARA, M., AZUMA, K., KAWAHARA, A., BASAKI, Y., HATTORI, S., YANAGAWA, T., TERAZAKI, Y., TAKAMORI, S., SHIROUZU, K., AIZAWA, H., NAKANO, K., KAGE, M., KUWANO, M. & ONO, M. 2009. Nuclear Y-Box binding protein-1, a predictive marker of prognosis, is correlated with expression of HER2/ErbB2 and HER3/ErbB3 in non-small cell lung cancer. *Journal of Thoracic Oncology*, 4, 1066-74.

- KASHIWAKURA, Y., TAMAYOSE, K., IWABUCHI, K., HIRAI, Y., SHIMADA, T., MATSUMOTO, K., NAKAMURA, T., WATANABE, M., OSHIMI, K. & DAIDA, H. 2005. Hepatocyte growth factor receptor is a coreceptor for adeno-associated virus type 2 infection. *J Virol*, 79, 609-14.
- KASPAR, B. K., ROTH, D. M., LAI, N. C., DRUMM, J. D., ERICKSON, D. A., MCKIRNAN, M. D. & HAMMOND, H. K. 2005. Myocardial gene transfer and long-term expression following intracoronary delivery of adeno-associated virus. *J Gene Med*, 7, 316-24.
- KATO, J., MATSUSHIME, H., HIEBERT, S. W., EWEN, M. E. & SHERR, C. J. 1993. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev*, 7, 331-42.
- KATZ, M. G., FARGNOLI, A. S., WILLIAMS, R. D., STEUERWALD, N. M., ISIDRO, A., IVANINA, A. V., SOKOLOVA, I. M. & BRIDGES, C. R. 2014. Safety and efficacy of high-dose adeno-associated virus 9 encoding sarcoplasmic reticulum Ca(2+) adenosine triphosphatase delivered by molecular cardiac surgery with recirculating delivery in ovine ischemic cardiomyopathy. *J Thorac Cardiovasc Surg*, 148, 1065-72.
- KAWAGUCHI, A., ASAKA, M. N., MATSUMOTO, K. & NAGATA, K. 2015a. Centrosome maturation requires YB-1 to regulate dynamic instability of microtubules for nucleus reassembly. *Sci Rep*, 5, 8768.
- KAWAGUCHI, A., HIROHAMA, M., HARADA, Y., OSARI, S. & NAGATA, K. 2015b. Influenza virus induces cholesterol-enriched endocytic recycling compartments for budzone formation via cell cycle-independent centrosome maturation. *PLoS Pathog*, 11, e1005284.
- KAWAGUCHI, A., MATSUMOTO, K. & NAGATA, K. 2012. YB-1 functions as a porter to lead influenza virus ribonucleoprotein complexes to microtubules. *J Virol*, 86, 11086-95.
- KELICH, J. M., MA, J., DONG, B., WANG, Q., CHIN, M., MAGURA, C. M., XIAO, W. & YANG, W. 2015. Super-resolution imaging of nuclear import of adeno-associated virus in live cells. *Molecular Therapy. Methods & Clinical Development*, 2, 15047.
- KELLEY, R. A., CONLEY, S. M., MAKKIA, R., WATSON, J. N., HAN, Z., COOPER, M. J. & NAASH, M. I. 2018. DNA nanoparticles are safe and nontoxic in non-human primate eyes. *Int J Nanomedicine*, 13, 1361-1379.
- KERN, A., SCHMIDT, K., LEDER, C., MULLER, O. J., WOBUS, C. E., BETTINGER, K., VON DER LIETH, C. W., KING, J. A. & KLEINSCHMIDT, J. A. 2003. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *Journal of Virology*, 77, 11072-81.
- KERR, D., CHANG, C. F., CHEN, N., GALLIA, G., RAJ, G., SCHWARTZ, B. & KHALILI, K. 1994. Transcription of a human neurotropic virus promoter in glial cells: effect of YB-1 on expression of the JC virus late gene. *J Virol*, 68, 7637-43.
- KHANDELWAL, P., PADALA, M. K., COX, J. & GUNTAKA, R. V. 2009. The N-terminal domain of y-box binding protein-1 induces cell cycle arrest in G2/M phase by binding to cyclin D1. *Int J Cell Biol*, 2009, 243532.
- KIM, B. T., KITAGAWA, H., TANAKA, J., TAMURA, J. & SUGAHARA, K. 2003. In vitro heparan sulfate polymerization: crucial roles of core protein moieties of primer substrates in addition to the EXT1-EXT2 interaction. *J Biol Chem*, 278, 41618-23.
- KIM, D., BAE, S., PARK, J., KIM, E., KIM, S., YU, H. R., HWANG, J., KIM, J. I. & KIM, J. S. 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*, 12, 237-43.
- KIM, E. L., WUSTENBERG, R., RUBSAM, A., SCHMITZ-SALUE, C., WARNECKE, G., BUCKER, E. M., PETTKUS, N., SPEIDEL, D., ROHDE, V., SCHULZ-SCHAEFFER, W., DEPPERT, W. & GIESE, A. 2010. Chloroquine activates the p53 pathway and induces apoptosis in human glioma cells. *Neuro Oncol*, 12, 389-400.
- KIM, E. R., SELJUTINA, A. A., BULDAKOV, I. A., EVDOKIMOVA, V., OVCHINNIKOV, L. P. & SOROKIN, A. V. 2013. The proteolytic YB-1 fragment interacts with DNA repair

- machinery and enhances survival during DNA damaging stress. *Cell Cycle*, 12, 3791-803.
- KIM, H. K., SONG, M., LEE, J., MENON, A. V., JUNG, S., KANG, Y. M., CHOI, J. W., WOO, E., KOH, H. C., NAM, J. W. & KIM, H. 2017. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods*, 14, 153-9.
- KIM, J. H., CHOI, E. Y., JUNG, E.-S., KWON, Y., LEE, D. S., HWANG, D. Y. & HWANG, E. S. 2009. Characterization of clones of human cell line infected with porcine endogenous retrovirus (PERV) from porcine cell line, PK-15. *Infection and Chemotherapy*, 41, 1.
- KING, J. A., DUBIELZIG, R., GRIMM, D. & KLEINSCHMIDT, J. A. 2001. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *Embo j*, 20, 3282-91.
- KITTS, P. A., AYRES, M. D. & POSSEE, R. D. 1990. Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vectors. *Nucleic Acids Res*, 18, 5667-72.
- KITTS, P. A. & POSSEE, R. D. 1993. A method for producing recombinant baculovirus expression vectors at high frequency. *Biotechniques*, 14, 810-7.
- KLEINSTIVER, B. P., PATTANAYAK, V., PREW, M. S., TSAI, S. Q., NGUYEN, N. T., ZHENG, Z. & JOUNG, J. K. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529, 490-5.
- KLEINSTIVER, B. P., PREW, M. S., TSAI, S. Q., TOPKAR, V. V., NGUYEN, N. T., ZHENG, Z., GONZALES, A. P., LI, Z., PETERSON, R. T., YEH, J. R., ARYEE, M. J. & JOUNG, J. K. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523, 481-5.
- KLJASHTORNY, V., NIKONOV, S., OVCHINNIKOV, L., LYABIN, D., VODOVAR, N., CURMI, P. & MANIVET, P. 2015. The cold shock domain of YB-1 segregates RNA from DNA by non-bonded interactions. *PLoS One*, 10, e0130318.
- KLOKS, C. P., SPRONK, C. A., LASONDER, E., HOFFMANN, A., VUISTER, G. W., GRZESIEK, S. & HILBERS, C. W. 2002. The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1. *J Mol Biol*, 316, 317-26.
- KOCH, M., LAUB, F., ZHOU, P., HAHN, R. A., TANAKA, S., BURGESSON, R. E., GERECKE, D. R., RAMIREZ, F. & GORDON, M. K. 2003. Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone. *J Biol Chem*, 278, 43236-44.
- KOHLBRENNER, E., ASLANIDI, G., NASH, K., SHKLYAEV, S., CAMPBELL-THOMPSON, M., BYRNE, B. J., SNYDER, R. O., MUZYCZKA, N., WARRINGTON, K. H. & ZOLOTUKHIN, S. 2005. Successful production of pseudotyped rAAV vectors using a modified baculovirus expression system. *Mol Ther*, 12, 1217-25.
- KOIKE, K., UCHIUMI, T., OHGA, T., TOH, S., WADA, M., KOHNO, K. & KUWANO, M. 1997. Nuclear translocation of the Y-box binding protein by ultraviolet irradiation. *FEBS Lett*, 417, 390-4.
- KOTAKE, Y., OZAWA, Y., HARADA, M., KITAGAWA, K., NIIDA, H., MORITA, Y., TANAKA, K., SUDA, T. & KITAGAWA, M. 2013. YB1 binds to and represses the p16 tumor suppressor gene. *Genes Cells*, 18, 999-1006.
- KOTIN, R. M., SINISCALCO, M., SAMULSKI, R. J., ZHU, X. D., HUNTER, L., LAUGHLIN, C. A., MCLAUGHLIN, S., MUZYCZKA, N., ROCCHI, M. & BERNS, K. I. 1990. Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A*, 87, 2211-5.
- KOTTERMAN, M. A. & SCHAFFER, D. V. 2014. Engineering adeno-associated viruses for clinical gene therapy. *Nat Rev Genet*, 15, 445-51.
- KOTTERMAN, M. A., YIN, L., STRAZZERI, J. M., FLANNERY, J. G., MERIGAN, W. H. & SCHAFFER, D. V. 2015. Antibody neutralization poses a barrier to intravitreal adeno-associated viral vector gene delivery to non-human primates. *Gene Ther*, 22, 116-26.

- KRÄTZER, F., ROSORIUS, O., HEGER, P., HIRSCHMANN, N., DOBNER, T., HAUBER, J. & STAUBER, R. H. 2000. The adenovirus type 5 E1B-55K oncoprotein is a highly active shuttle protein and shuttling is independent of E4orf6, p53 and Mdm2. *Oncogene*, 19, 850-7.
- KUBOTA, H., HAYASHI, Y., KUBOTA, Y., COWARD, K. & PARRINGTON, J. 2005. Comparison of two methods of in vivo gene transfer by electroporation. *Fertil Steril*, 83 Suppl 1, 1310-8.
- KUMAR, A. & PEDERSON, T. 1975. Comparison of proteins bound to heterogeneous nuclear RNA and messenger RNA in HeLa cells. *J Mol Biol*, 96, 353-65.
- KUMAR, S., STECHER, G., LI, M., KNYAZ, C. & TAMURA, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 35, 1547-9.
- KYTE, J. & DOOLITTLE, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157, 105-32.
- LABOW, M. A. & BERNS, K. I. 1988. The adeno-associated virus rep gene inhibits replication of an adeno-associated virus/simian virus 40 hybrid genome in cos-7 cells. *J Virol*, 62, 1705-12.
- LABRIE, S. J., SAMSON, J. E. & MOINEAU, S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol*, 8, 317-27.
- LABUHN, M., ADAMS, F. F., NG, M., KNOESS, S., SCHAMBACH, A., CHARPENTIER, E. M., SCHWARZER, A., MATEO, J. L., KLUSMANN, J. H. & HECKL, D. 2018. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res*, 46, 1375-85.
- LADOMERY, M. & SOMMERVILLE, J. 1994. Binding of Y-box proteins to RNA: involvement of different protein domains. *Nucleic Acids Research*, 22, 5582-9.
- LAGUETTE, N., RAHM, N., SOBHIAN, B., CHABLE-BESSIA, C., MUNCH, J., SNOECK, J., SAUTER, D., SWITZER, W. M., HENEINE, W., KIRCHHOFF, F., DELSUC, F., TELENTI, A. & BENKIRANE, M. 2012. Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe*, 11, 205-17.
- LAI, Y. K., HSU, J. T., CHU, C. C., CHANG, T. Y., PAN, K. L. & LIN, C. C. 2012. Enhanced recombinant protein production and differential expression of molecular chaperones in sf-caspase-1-repressed stable cells after baculovirus infection. *BMC Biotechnol*, 12, 83.
- LAKHTER, A. J., SAHU, R. P., SUN, Y., KAUFMANN, W. K., ANDROPHY, E. J., TRAVERS, J. B. & NAIDU, S. R. 2013. Chloroquine promotes apoptosis in melanoma cells by inhibiting BH3 domain-mediated PUMA degradation. *J Invest Dermatol*, 133, 2247-54.
- LALONDE, S., STONE, O. A., LESSARD, S., LAVERTU, A., DESJARDINS, J., BEAUDOIN, M., RIVAS, M., STAINIER, D. Y. R. & LETTRE, G. 2017. Frameshift indels introduced by genome editing can lead to in-frame exon skipping. *PLoS One*, 12, e0178700.
- LAMAZE, C., CHUANG, T. H., TERLECKY, L. J., BOKOCH, G. M. & SCHMID, S. L. 1996. Regulation of receptor-mediated endocytosis by Rho and Rac. *Nature*, 382, 177-9.
- LANDSMAN, D. 1992. RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res*, 20, 2861-4.
- LANGEREIS, M. A., ROSAS-ACOSTA, G., MULDER, K. & WILSON, V. G. 2007. Production of sumoylated proteins using a baculovirus expression system. *J Virol Methods*, 139, 189-94.
- LASHAM, A., LINDRIDGE, E., RUDERT, F., ONRUST, R. & WATSON, J. 2000. Regulation of the human fas promoter by YB-1, Puralpha and AP-1 transcription factors. *Gene*, 252, 1-13.
- LASHAM, A., MOLONEY, S., HALE, T., HOMER, C., ZHANG, Y. F., MURISON, J. G., BRAITHWAITE, A. W. & WATSON, J. 2003. The Y-box-binding protein, YB1, is a potential negative regulator of the p53 tumor suppressor. *J Biol Chem*, 278, 35516-23.

- LAURING, A. S., FRYDMAN, J. & ANDINO, R. 2013. The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol*, 11, 327-36.
- LEBBINK, R. J., DE JONG, D. C., WOLTERS, F., KRUSE, E. M., VAN HAM, P. M., WIERTZ, E. J. & NIJHUIS, M. 2017. A combinational CRISPR/Cas9 gene-editing approach can halt HIV replication and prevent viral escape. *Sci Rep*, 7, 41968.
- LEE, J. K., JEONG, E., LEE, J., JUNG, M., SHIN, E., KIM, Y.-H., LEE, K., KIM, D., KIM, J.-S. & KIM, S. 2017a. Directed evolution of CRISPR-Cas9 to increase its specificity. *bioRxiv* 237040.
- LEE, K., CONBOY, M., PARK, H. M., JIANG, F., KIM, H. J., DEWITT, M. A., MACKLEY, V. A., CHANG, K., RAO, A., SKINNER, C., SHOBHA, T., MEHDIPOUR, M., LIU, H., HUANG, W.-C., LAN, F., BRAY, N. L., LI, S., CORN, J. E., KATAOKA, K., DOUDNA, J. A., CONBOY, I. & MURTHY, N. 2017b. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nature Biomedical Engineering*, 1, 889-901.
- LEFESVRE, P., ATTEMA, J. & VAN BEKKUM, D. 2002. A comparison of efficacy and toxicity between electroporation and adenoviral gene transfer. *BMC Mol Biol*, 3, 12.
- LESSARD, S., FRANCIOLI, L., ALFOLDI, J., TARDIF, J. C., ELLINOR, P. T., MACARTHUR, D. G., LETTRE, G., ORKIN, S. H. & CANVER, M. C. 2017. Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proc Natl Acad Sci U S A*, 114, E11257-66.
- LEVASSEUR, A., BEKLIZ, M., CHABRIERE, E., PONTAROTTI, P., LA SCOLA, B. & RAOULT, D. 2016. MIMIVIRE is a defence system in mimivirus that confers resistance to viroplasm. *Nature*, 531, 249-52.
- LEVY, A., GOREN, M. G., YOSEF, I., AUSTER, O., MANOR, M., AMITAI, G., EDGAR, R., QIMRON, U. & SOREK, R. 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, 520, 505-10.
- LEVY, M., BUSKILA, D., GLADMAN, D. D., UROWITZ, M. B. & KOREN, G. 1991. Pregnancy outcome following first trimester exposure to chloroquine. *Am J Perinatol*, 8, 174-8.
- LI, E., STUPACK, D., BOKOCH, G. M. & NEMEROW, G. R. 1998. Adenovirus endocytosis requires actin cytoskeleton reorganization mediated by Rho family GTPases. *J Virol*, 72, 8806-12.
- LI, H., SHENG, C., WANG, S., YANG, L., LIANG, Y., HUANG, Y., LIU, H., LI, P., YANG, C., YANG, X., JIA, L., XIE, J., WANG, L., HAO, R., DU, X., XU, D., ZHOU, J., LI, M., SUN, Y., TONG, Y., LI, Q., QIU, S. & SONG, H. 2017. Removal of integrated hepatitis B virus DNA using CRISPR-Cas9. *Front Cell Infect Microbiol*, 7, 91.
- LI, J., SAMULSKI, R. J. & XIAO, X. 1997. Role for highly regulated rep gene expression in adeno-associated virus vector production. *J Virol*, 71, 5236-43.
- LI, M., LIU, H., HAN, J., LIU, J., WANG, R., ZHAO, D., ZHOU, J. & XIANG, H. 2013. Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon *Haloferax mediterranei*. *J Bacteriol*, 195, 867-75.
- LI, S. & HU, G. F. 2012. Emerging role of angiogenin in stress response and cell survival under adverse conditions. *J Cell Physiol*, 227, 2822-6.
- LI, S., ZHANG, Q. N., ZHANG, X. T., ZHENG, X. Y., LV, Y. F. & HAO, Z. M. 2014. Cytomegalovirus immediate-early promoter efficiently drives heterogeneous gene expression in *Spodoptera frugiperda* (Sf9) insect cells. *Cell Mol Biol (Noisy-le-grand)*, 60, 6-11.
- LI, W., WANG, X. & GAO, G. 2012. Expression of YB-1 enhances production of murine leukemia virus vectors by stabilizing genomic viral RNA. *Protein Cell*, 3, 943-9.
- LI, Y., PAN, S., ZHANG, Y., REN, M., FENG, M., PENG, N., CHEN, L., LIANG, Y. X. & SHE, Q. 2016. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res*, 44, e34.

- LI, Z., DAY, C. P., YANG, J. Y., TSAI, W. B., LOZANO, G., SHIH, H. M. & HUNG, M. C. 2004. Adenoviral E1A targets Mdm4 to stabilize tumor suppressor p53. *Cancer Res*, 64, 9080-5.
- LI, Z., WEI, J., FAN, Y., MEI, X., HE, Q., ZHANG, Y., LI, T., LONG, M., CHEN, J., LIU, T.-B., BAO, J., XIANG, Z., PAN, G., LI, C. & ZHOU, Z. 2018.
- LIANG, X., POTTER, J., KUMAR, S., ZOU, Y., QUINTANILLA, R., SRIDHARAN, M., CARTE, J., CHEN, W., ROARK, N., RANGANATHAN, S., RAVINDER, N. & CHESNUT, J. D. 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol*, 208, 44-53.
- LIAO, H. K., GU, Y., DIAZ, A., MARLETT, J., TAKAHASHI, Y., LI, M., SUZUKI, K., XU, R., HISHIDA, T., CHANG, C. J., ESTEBAN, C. R., YOUNG, J. & IZPISUA BELMONTE, J. C. 2015. Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells. *Nat Commun*, 6, 6413.
- LILYESTROM, W., KLEIN, M. G., ZHANG, R., JOACHIMIAK, A. & CHEN, X. S. 2006. Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev*, 20, 2373-82.
- LIM, Y., BAK, S. Y., SUNG, K., JEONG, E., LEE, S. H., KIM, J. S., BAE, S. & KIM, S. K. 2016. Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat Commun*, 7.
- LIN, J. Y. & SIMMONS, D. T. 1991. The ability of large T antigen to complex with p53 is necessary for the increased life span and partial transformation of human cells by simian virus 40. *J Virol*, 65, 6447-53.
- LIN, S. R., YANG, H. C., KUO, Y. T., LIU, C. J., YANG, T. Y., SUNG, K. C., LIN, Y. Y., WANG, H. Y., WANG, C. C., SHEN, Y. C., WU, F. Y., KAO, J. H., CHEN, D. S. & CHEN, P. J. 2014. The CRISPR/Cas9 System Facilitates Clearance of the Intrahepatic HBV Templates In Vivo. *Mol Ther Nucleic Acids*, 3, e186.
- LINDEN, R. M., WINOCOUR, E. & BERNS, K. I. 1996. The recombination signals for Adeno-associated virus site-specific integration. *Proc Natl Acad Sci U S A*, 93, 7966-72.
- LING, C., LI, B., MA, W. & SRIVASTAVA, A. 2016. Development of optimized AAV serotype vectors for high-efficiency transduction at further reduced doses. *Hum Gene Ther Methods*, 27, 143-9.
- LIU, C., DALBY, B., CHEN, W., KILZER, J. M. & CHIOU, H. C. 2008. Transient transfection factors for high-level recombinant protein production in suspension cultured mammalian cells. *Mol Biotechnol*, 39, 141-53.
- LIU, R., PAXTON, W. A., CHOE, S., CERADINI, D., MARTIN, S. R., HORUK, R., MACDONALD, M. E., STUHLMANN, H., KOUP, R. A. & LANDAU, N. R. 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*, 86, 367-77.
- LIU, T. Y., IAVARONE, A. T. & DOUDNA, J. A. 2017. RNA and DNA targeting by a reconstituted *Thermus thermophilus* type III-A CRISPR-Cas system. *PLoS One*, 12, e0170552.
- LIU, Y., JOO, K. I. & WANG, P. 2013. Endocytic processing of adeno-associated virus type 8 vectors for transduction of target cells. *Gene Ther*, 20, 308-17.
- LOCK, M., ALVIRA, M., VANDENBERGHE, L. H., SAMANTA, A., TOELEN, J., DEBYSER, Z. & WILSON, J. M. 2010. Rapid, simple, and versatile manufacturing of recombinant adeno-associated viral vectors at scale. *Hum Gene Ther*, 21, 1259-71.
- LONG, C., AMOASII, L., MIREAULT, A. A., MCANALLY, J. R., LI, H., SANCHEZ-ORTIZ, E., BHATTACHARYYA, S., SHELTON, J. M., BASSEL-DUBY, R. & OLSON, E. N. 2016. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science*, 351, 400-3.
- LOPEZ-VIDAL, J., GOMEZ-SEBASTIAN, S., BARCENA, J., NUNEZ MDEL, C., MARTINEZ-ALONSO, D., DUDOGNON, B., GUIJARRO, E. & ESCRIBANO, J. M. 2015. Improved production

- efficiency of virus-like particles by the baculovirus expression vector system. *PLoS One*, 10, e0140039.
- LOWE, S. W. & RULEY, H. E. 1993. Stabilization of the p53 tumor suppressor is induced by adenovirus 5 E1A and accompanies apoptosis. *J Virol*, 7, 4.
- LU, Z. H., BOOKS, J. T. & LEY, T. J. 2005. YB-1 is important for late-stage embryonic development, optimal cellular stress responses, and the prevention of premature senescence. *Mol Cell Biol*, 25, 4625-37.
- LU, Z. H., BOOKS, J. T. & LEY, T. J. 2006. Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. *Mol Cell Biol*, 26, 8410-7.
- LUCKOW, V. A., LEE, S. C., BARRY, G. F. & OLINS, P. O. 1993. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J Virol*, 67, 4566-79.
- LUO, K., EHRLICH, E., XIAO, Z., ZHANG, W., KETNER, G. & YU, X. F. 2007. Adenovirus E4orf6 assembles with Cullin5-ElonginB-ElonginC E3 ubiquitin ligase through an HIV/SIV Vif-like BC-box to regulate p53. *FASEB J*, 21, 1742-50.
- LUSBY, E., BOHENZKY, R. & BERNS, K. I. 1981. Inverted terminal repetition in Adeno-associated virus DNA: independence of the orientation at either end of the genome. *J Virol*, 37, 1083-6.
- LUTHMAN, H. & MAGNUSSON, G. 1983. High efficiency polyoma DNA transfection of chloroquine treated cells. *Nucleic Acids Res*, 11, 1295-308.
- LUTZ, M., WEMPE, F., BAHR, I., ZOPF, D. & VON MELCHNER, H. 2006. Proteasomal degradation of the multifunctional regulator YB-1 is mediated by an F-Box protein induced during programmed cell death. *FEBS Lett*, 580, 3921-30.
- LUX, K., GOERLITZ, N., SCHLEMMINGER, S., PERABO, L., GOLDNAU, D., ENDELL, J., LEIKE, K., KOFLER, D. M., FINKE, S., HALLEK, M. & BUNING, H. 2005. Green fluorescent protein-tagged adeno-associated virus particles allow the study of cytosolic and nuclear trafficking. *J Virol*, 79, 11776-87.
- LYABIN, D. N., ELISEEVA, I. A. & OVCHINNIKOV, L. P. 2014. YB-1 protein: functions and regulation. *Wiley Interdiscip Rev RNA*, 5, 95-110.
- LYONS, S. M., ACHORN, C., KEDERSHA, N. L., ANDERSON, P. J. & IVANOV, P. 2016. YB-1 regulates tRNA-induced Stress Granule formation but not translational repression. *Nucleic Acids Res*, 44, 6949-60.
- MAEDA, T., KUSAKABE, T., LEE, J. M., MIYAGAWA, Y., KOGA, K. & KAWAGUCHI, Y. 2005. Efficient nonviral gene transfer mediated by polyethylenimine in an insect cell line. *J Insect Biotechnol and Sericol*, 74, 21-6.
- MAEDER, M. L., LINDER, S. J., CASCIO, V. M., FU, Y., HO, Q. H. & JOUNG, J. K. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*, 10, 977-9.
- MAHON, M. J. 2011. Vectors bicistronically linking a gene of interest to the SV40 large T antigen in combination with the SV40 origin of replication enhance transient protein expression and luciferase reporter activity. *Biotechniques*, 51, 119-28.
- MAIOROV, V. N. & CRIPPEN, G. M. 1994. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol*, 235, 625-34.
- MAKAROVA, K. S., ARAVIND, L., GRISHIN, N. V., ROGOZIN, I. B. & KOONIN, E. V. 2002. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res*, 30, 482-96.
- MAKAROVA, K. S., ARAVIND, L., WOLF, Y. I. & KOONIN, E. V. 2011a. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct*, 6, 38.
- MAKAROVA, K. S., HAFT, D. H., BARRANGOU, R., BROUNS, S. J., CHARPENTIER, E., HORVATH, P., MOINEAU, S., MOJICA, F. J., WOLF, Y. I., YAKUNIN, A. F., VAN DER OOST, J. &

- KOONIN, E. V. 2011b. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*, 9, 467-77.
- MAKAROVA, K. S., WOLF, Y. I., ALKHNABASHI, O. S., COSTA, F., SHAH, S. A., SAUNDERS, S. J., BARRANGOU, R., BROUNS, S. J., CHARPENTIER, E., HAFT, D. H., HORVATH, P., MOINEAU, S., MOJICA, F. J., TERNS, R. M., TERNS, M. P., WHITE, M. F., YAKUNIN, A. F., GARRETT, R. A., VAN DER OOST, J., BACKOFEN, R. & KOONIN, E. V. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, 13, 722-36.
- MALI, P., AACH, J., STRANGES, P. B., ESVELT, K. M., MOOSBURNER, M., KOSURI, S., YANG, L. & CHURCH, G. M. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*, 31, 833-8.
- MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013b. RNA-guided human genome engineering via Cas9. *Science*, 339, 823-6.
- MANI, A., YADAVA, P. K. & GUPTA, D. K. 2012. Cold shock domain protein from *Philosamia ricini* prefers single-stranded nucleic acids binding. *J Biomol Struct Dyn*, 30, 532-41.
- MANO, M., IPPODRINO, R., ZENTILIN, L., ZACCHIGNA, S. & GIACCA, M. 2015. Genome-wide RNAi screening identifies host restriction factors critical for in vivo AAV transduction. *Proc Natl Acad Sci U S A*, 112, 11276-81.
- MANSOURI, M., BELLON-ECHEVERRIA, I., RIZK, A., EHSAEI, Z., CIANCIOLO COSENTINO, C., SILVA, C. S., XIE, Y., BOYCE, F. M., DAVIS, M. W., NEUHAUSS, S. C., TAYLOR, V., BALLMER-HOFER, K., BERGER, I. & BERGER, P. 2016. Highly efficient baculovirus-mediated multigene delivery in primary cells. *Nat Commun*, 7, 11529.
- MARTIN, M., HUA, L., WANG, B., WEI, H., PRABHU, L., HARTLEY, A. V., JIANG, G., LIU, Y. & LU, T. 2017. Novel serine 176 phosphorylation of YBX1 activates NF-kappaB in colon cancer. *J Biol Chem*, 292, 3433-44.
- MARTINEZ-SOLIS, M., GOMEZ-SEBASTIAN, S., ESCRIBANO, J. M., JAKUBOWSKA, A. K. & HERRERO, S. 2016. A novel baculovirus-derived promoter with high activity in the baculovirus expression system. *PeerJ*, 4, e2183.
- MARTINEZ, J., PATKANIEWSKA, A., URLAUB, H., LUHRMANN, R. & TUSCHL, T. 2002. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, 110, 563-74.
- MARTINO, A. T., BASNER-TSCHAKARJAN, E., MARKUSIC, D. M., FINN, J. D., HINDERER, C., ZHOU, S., OSTROV, D. A., SRIVASTAVA, A., ERTL, H. C., TERHORST, C., HIGH, K. A., MINGOZZI, F. & HERZOG, R. W. 2013. Engineered AAV vector minimizes in vivo targeting of transduced hepatocytes by capsid-specific CD8+ T cells. *Blood*, 121, 2224-33.
- MASUDA, T., WAN, J., YERRABELLI, A., BERLINICKE, C., KALLMAN, A., QIAN, J. & ZACK, D. J. 2016. Off Target, but Sequence-Specific, shRNA-Associated Trans-Activation of Promoter Reporters in Transient Transfection Assays. *PLoS One*, 11, e0167867.
- MASUMOTO, M., OHDE, T., SHIOMI, K., YAGINUMA, T. & NIIMI, T. 2012. A baculovirus immediate-early gene, ie1, promoter drives efficient expression of a transgene in both *Drosophila melanogaster* and *Bombyx mori*. *PLoS ONE*, 7, e49323.
- MATSUMOTO, S., UCHIUMI, T., TANAMACHI, H., SAITO, T., YAGI, M., TAKAZAKI, S., KANKI, T. & KANG, D. 2012. Ribonucleoprotein Y-box-binding protein-1 regulates mitochondrial oxidative phosphorylation (OXPHOS) protein expression after serum stimulation through binding to OXPHOS mRNA. *Biochem J*, 443, 573-84.
- MATSUSHITA, T., ELLIGER, S., ELLIGER, C., PODSAKOFF, G., VILLARREAL, L., KURTZMAN, G. J., IWAKI, Y. & COLOSI, P. 1998. Adeno-associated virus vectors can be efficiently produced without helper virus. *Gene Ther*, 5, 938-45.
- MCCARTY, D. M., PEREIRA, D. J., ZOLOTUKHIN, I., ZHOU, X., RYAN, J. H. & MUZYCZKA, N. 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. *J Virol*, 68, 4988-97.

- MCGUFFIN, L. J. 2008. Intrinsic disorder prediction from the analysis of multiple protein fold recognition models. *Bioinformatics*, 24, 1798-804.
- MCGUFFIN, L. J., ATKINS, J. D., SALEHE, B. R., SHUID, A. N. & ROCHE, D. B. 2015. IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Res*, 43, W169-73.
- MCGUFFIN, L. J., SHUID, A. N., KEMPSTER, R., MAGHRABI, A. H. A., NEALON, J. O., SALEHE, B. R., ATKINS, J. D. & ROCHE, D. B. 2018. Accurate template-based modeling in CASP12 using the IntFOLD4-TS, ModFOLD6, and ReFOLD methods. *Proteins*, 86 Suppl 1, 335-344.
- MEDINA, E., GOLDMANN, O., TOPPEL, A. W. & CHHATWAL, G. S. 2003. Survival of *Streptococcus pyogenes* within host phagocytic cells: a pathogenic mechanism for persistence and systemic invasion. *J Infect Dis*, 187, 597-603.
- MEGHROUS, J., AUCOIN, M. G., JACOB, D., CHAHAL, P. S., ARCAND, N. & KAMEN, A. A. 2005. Production of recombinant adeno-associated viral vectors using a baculovirus/insect cell suspension culture system: from shake flasks to a 20-L bioreactor. *Biotechnol Prog*, 21, 154-60.
- MEISSNER, P., PICK, H., KULANGARA, A., CHATELLARD, P., FRIEDRICH, K. & WURM, F. M. 2001. Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnol Bioeng*, 75, 197-203.
- MEKLER, V., MINAKHIN, L. & SEVERINOV, K. 2017. Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc Natl Acad Sci U S A*, 114, 5443-8.
- MIESBACH, W., MEIJER, K., COPPENS, M., KAMPMANN, P., KLAMROTH, R., SCHUTGENS, R., TANGELDER, M., CASTAMAN, G., SCHWABLE, J., BONIG, H., SEIFRIED, E., CATTANEO, F., MEYER, C. & LEEBEEK, F. W. G. 2018. Gene therapy with Adeno-associated virus vector 5-human factor IX in adults with hemophilia B. *Blood*, 131, 1022-31.
- MIETZ, J. A., UNGER, T., HUIBREGTSE, J. M. & HOWLEY, P. M. 1992. The transcriptional transactivation function of wild-type p53 is inhibited by SV40 large T-antigen and by HPV-16 E6 oncoprotein. *Embo j*, 11, 5013-20.
- MIETZSCH, M., GRASSE, S., ZURAWSKI, C., WEGER, S., BENNETT, A., AGBANDJE-MCKENNA, M., MUZYCZKA, N., ZOLOTUKHIN, S. & HEILBRONN, R. 2014. OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1-12 vectors for gene therapy. *Hum Gene Ther*, 25, 212-22.
- MINGOZZI, F., ANGUOLA, X. M., PAVANI, G., CHEN, Y., DAVIDSON, R. J., HUI, D. J., YAZICIOGLU, M., ELKOUBY, L., HINDERER, C. J., FAELLA, A., HOWARD, C., TAI, A., PODSAKOFF, G. M., ZHOU, S., BASNER-TSCHAKARJAN, E., WRIGHT, J. F. & HIGH, K. A. 2013. Overcoming preexisting humoral immunity to AAV using capsid decoys. *Sci Transl Med*, 5, 194ra92.
- MINICH, W. B., MAIDEBURA, I. P. & OVCHINNIKOV, L. P. 1993. Purification and characterization of the major 50-kDa repressor protein from cytoplasmic mRNP of rabbit reticulocytes. *Eur J Biochem*, 212, 633-8.
- MIRANDA, E. A., DE-MURCIA, G. & MURCIA, J. 1997. Large-scale production and purification of recombinant protein from an insect cell/baculovirus system in Erlenmeyer flasks: Application to the chicken poly(ADP-ribose) polymerase catalytic domain. *Braz J Med Biol Res*, 30, 923-8.
- MITAL, A., KUMARI, D., GUPTA, M. & GOYLE, S. 1998. Molecular characterisation of Duchenne muscular dystrophy and phenotypic correlation. *J Neurol Sci*, 157, 179-86.
- MITCHELL, A. M., HIRSCH, M. L., LI, C. & SAMULSKI, R. J. 2014. Promyelocytic leukemia protein is a cell-intrinsic factor inhibiting parvovirus DNA replication. *J Virol*, 88, 925-36.
- MOGHIMI, S. M., SYMONDS, P., MURRAY, J. C., HUNTER, A. C., DEBSKA, G. & SZEWCZYK, A. 2005. A two-stage poly(ethylenimine)-mediated cytotoxicity: implications for gene transfer/therapy. *Mol Ther*, 11, 990-5.

- MOHAMADIPOOR, M., HABIBI, R. M., MASROORI, N., MOHAMMADI, R. A. & SAKI, S. 2009. Expression of Green Fluorescent Protein (GFP) using In Vitro translation cell free system. *DARU*, 17, 60-3.
- MOHSENI, N., JAHANIAN-NAJAFABADI, A., KAZEMI-LOMEDASHT, F., AREZOMAND, R., HABIBI-ANBOUHI, M., SHAHBAZZADEH, D. & BEHDANI, M. 2016. Recombinant expression and purification of functional vascular endothelial growth factor-121 in the baculovirus expression system. *Asian Pac J Trop Med*, 9, 1195-9.
- MOIANI, A., PALEARI, Y., SARTORI, D., MEZZADRA, R., MICCIO, A., CATTOGLIO, C., COCCHIARELLA, F., LIDONNICI, M. R., FERRARI, G. & MAVILIO, F. 2012. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest*, 122, 1653-66.
- MONTAGUE, T. G., CRUZ, J. M., GAGNON, J. A., CHURCH, G. M. & VALEN, E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res*, 42, W401-7.
- MORI, S., WANG, L., TAKEUCHI, T. & KANDA, T. 2004. Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein. *Virology*, 330, 375-83.
- MORTENSEN, R., NISSEN, T. N., BLAUENFELDT, T., CHRISTENSEN, J. P., ANDERSEN, P. & DIETRICH, J. 2015. Adaptive immunity against *Streptococcus pyogenes* in adults involves increased IFN-gamma and IgG3 responses compared with children. *J Immunol*, 195, 1657-64.
- MOSS, R. B., RODMAN, D., SPENCER, L. T., AITKEN, M. L., ZEITLIN, P. L., WALTZ, D., MILLA, C., BRODY, A. S., CLANCY, J. P., RAMSEY, B., HAMBLETT, N. & HEALD, A. E. 2004. Repeated adeno-associated virus serotype 2 aerosol-mediated cystic fibrosis transmembrane regulator gene transfer to the lungs of patients with cystic fibrosis: a multicenter, double-blind, placebo-controlled trial. *Chest*, 125, 509-21.
- MU, X., LI, W., WANG, X. & GAO, G. 2013. YB-1 stabilizes HIV-1 genomic RNA and enhances viral production. *Protein Cell*, 4, 591-7.
- MUKHERJEE, S., GHOSH, R. N. & MAXFIELD, F. R. 1997. Endocytosis. *Physiol Rev*, 77, 759-803.
- MULEPATI, S., HEROUX, A. & BAILEY, S. 2014. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, 345, 1479-84.
- MYERS, J. W., JONES, J. T., MEYER, T. & FERRELL, J. E., JR. 2003. Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nat Biotechnol*, 21, 324-8.
- MYERS, M. W. & CARTER, B. J. 1980. Assembly of adeno-associated virus. *Virology*, 102, 71-82.
- NAKAMINAMI, K., KARLSON, D. T. & IMAI, R. 2006. Functional conservation of cold shock domains in bacteria and higher plants. *Proc Natl Acad Sci U S A*, 103, 10122-7.
- NANDAKUMAR, S., MA, H. & KHAN, A. S. 2017. Whole-Genome Sequence of the *Spodoptera frugiperda* Sf9 Insect Cell Line. *Genome Announc*, 5.
- NARASIMHA, A. M., KAULICH, M., SHAPIRO, G. S., CHOI, Y. J., SICINSKI, P. & DOWDY, S. F. 2014. Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. *Elife*, 3.
- NASH, K., CHEN, W., SALGANIK, M. & MUZYCZKA, N. 2009. Identification of cellular proteins that interact with the adeno-associated virus rep protein. *J Virol*, 83, 454-69.
- NATHWANI, A. C., GRAY, J. T., MCINTOSH, J., NG, C. Y., ZHOU, J., SPENCE, Y., COCHRANE, M., GRAY, E., TUDDENHAM, E. G. & DAVIDOFF, A. M. 2007. Safe and efficient transduction of the liver after peripheral vein infusion of self-complementary AAV vector results in stable therapeutic expression of human FIX in nonhuman primates. *Blood*, 109, 1414-21.
- NATHWANI, A. C., REISS, U. M., TUDDENHAM, E. G., ROSALES, C., CHOWDARY, P., MCINTOSH, J., DELLA PERUTA, M., LHERITEAU, E., PATEL, N., RAJ, D., RIDDELL, A., PIE, J., RANGARAJAN, S., BEVAN, D., RECHT, M., SHEN, Y. M., HALKA, K. G., BASNER-TSCHAKARJAN, E., MINGOZZI, F., HIGH, K. A., ALLAY, J., KAY, M. A., NG, C. Y., ZHOU, J.,

- CANCIO, M., MORTON, C. L., GRAY, J. T., SRIVASTAVA, D., NIENHUIS, A. W. & DAVIDOFF, A. M. 2014. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med*, 371, 1994-2004.
- NATHWANI, A. C., TUDDENHAM, E. G., RANGARAJAN, S., ROSALES, C., MCINTOSH, J., LINCH, D. C., CHOWDARY, P., RIDDELL, A., PIE, A. J., HARRINGTON, C., O'BEIRNE, J., SMITH, K., PASI, J., GLADER, B., RUSTAGI, P., NG, C. Y., KAY, M. A., ZHOU, J., SPENCE, Y., MORTON, C. L., ALLAY, J., COLEMAN, J., SLEEP, S., CUNNINGHAM, J. M., SRIVASTAVA, D., BASNER-TSCHAKARJAN, E., MINGOZZI, F., HIGH, K. A., GRAY, J. T., REISS, U. M., NIENHUIS, A. W. & DAVIDOFF, A. M. 2011. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med*, 365, 2357-65.
- NAULT, J. C., DATTA, S., IMBEAUD, S., FRANCONI, A., MALLET, M., COUCHY, G., LETOUZE, E., PILATI, C., VERRET, B., BLANC, J. F., BALABAUD, C., CALDERARO, J., LAURENT, A., LETEXIER, M., BIOULAC-SAGE, P., CALVO, F. & ZUCMAN-ROSSI, J. 2015. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet*, 47, 1187-93.
- NAUMER, M., SONNTAG, F., SCHMIDT, K., NIETO, K., PANKE, C., DAVEY, N. E., POPA-WAGNER, R. & KLEINSCHMIDT, J. A. 2012. Properties of the adeno-associated virus assembly-activating protein. *J Virol*, 86, 13038-48.
- NEEDLEMAN, S. B. & WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48, 443-53.
- NEGRETE, A., YANG, L. C., MENDEZ, A. F., LEVY, J. R. & KOTIN, R. M. 2007. Economized large-scale production of high yield of rAAV for gene therapy applications exploiting baculovirus expression system. *J Gene Med*, 9, 938-48.
- NEKRASOV, M. P., IVSHINA, M. P., CHERNOV, K. G., KOVRIGINA, E. A., EVDOKIMOVA, V. M., THOMAS, A. A., HERSHEY, J. W. & OVCHINNIKOV, L. P. 2003. The mRNA-binding protein YB-1 (p50) prevents association of the eukaryotic initiation factor eIF4G with mRNA and inhibits protein synthesis at the initiation stage. *J Biol Chem*, 278, 13936-43.
- NELSON, C. E., HAKIM, C. H., OUSTEROUT, D. G., THAKORE, P. I., MOREB, E. A., CASTELLANOS RIVERA, R. M., MADHAVAN, S., PAN, X., RAN, F. A., YAN, W. X., ASOKAN, A., ZHANG, F., DUAN, D. & GERSBACH, C. A. 2016. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*, 351, 403-7.
- NEUMANN, E., SCHAEFER-RIDDER, M., WANG, Y. & HOFSCHNEIDER, P. H. 1982. Gene transfer into mouse lymphoma cells by electroporation in high electric fields. *EMBO J*, 1, 841-5.
- NEVINS, J. R. 1990. Adenovirus E1A-dependent trans-activation of transcription. *Semin Cancer Biol*, 1, 59-68.
- NEWKIRK, K., Q. FENG, W., N. JIANG, W., TEJERO, R., EMERSON, S. D., INOUE, M. & MONTELIONE, G. 1994. Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA. *Proc Natl Acad Sci U S A*, 91, 5114-8.
- NICOLSON, S. C. & SAMULSKI, R. J. 2014. Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus. *J Virol*, 88, 4132-44.
- NISHIMASU, H., CONG, L., YAN, W. X., RAN, F. A., ZETSCHKE, B., LI, Y., KURABAYASHI, A., ISHITANI, R., ZHANG, F. & NUREKI, O. 2015. Crystal structure of *Staphylococcus aureus* Cas9. *Cell*, 162, 1113-26.
- NISHIMASU, H., RAN, F. A., HSU, P. D., KONERMANN, S., SHEHATA, S. I., DOHMAE, N., ISHITANI, R., ZHANG, F. & NUREKI, O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156, 935-49.
- NISHITA, Y. & TAKIYA, S. 2005. Gene expression profiling and molecular dissection of the *Bombyx* Y-box protein BYB. *J Insect Biotechnol Sericol*, 74, 85-93.
- NIU, D., WEI, H. J., LIN, L., GEORGE, H., WANG, T., LEE, I. H., ZHAO, H. Y., WANG, Y., KAN, Y., SHROCK, E., LESH, E., WANG, G., LUO, Y., QING, Y., JIAO, D., ZHAO, H., ZHOU, X.,

- WANG, S., WEI, H., GUELL, M., CHURCH, G. M. & YANG, L. 2017. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science*, 357, 1303-1307.
- NOBES, C. D. & HALL, A. 1995. Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell*, 81, 53-62.
- NONNENMACHER, M. & WEBER, T. 2011. Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway. *Cell Host Microbe*, 10, 563-76.
- NONNENMACHER, M. & WEBER, T. 2012. Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther*, 19, 649-58.
- NUNEZ, J. K., KRANZUSCH, P. J., NOESKE, J., WRIGHT, A. V., DAVIES, C. W. & DOUDNA, J. A. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol*, 21, 528-34.
- NUNEZ, J. K., LEE, A. S., ENGELMAN, A. & DOUDNA, J. A. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, 519, 193-8.
- NUSSBAUM, A. L., DAVOLI, D., GANEM, D. & FAREED, G. C. 1976. Construction and propagation of a defective simian virus 40 genome bearing an operator from bacteriophage lambda. *Proc Natl Acad Sci U S A*, 73, 1068-72.
- O'DONNELL, J., TAYLOR, K. A. & CHAPMAN, M. S. 2009. Adeno-associated virus-2 and its primary cellular receptor--Cryo-EM structure of a heparin complex. *Virology*, 385, 434-43.
- O'GEEN, H., HENRY, I. M., BHAKTA, M. S., MECKLER, J. F. & SEGAL, D. J. 2015. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res*, 43, 3389-404.
- OGAY, I. D., LIHORADOVA, O. A., AZIMOVA SH, S., ABDUKARIMOV, A. A., SLACK, J. M. & LYNN, D. E. 2006. Transfection of insect cell lines using polyethylenimine. *Cytotechnology*, 51, 89-98.
- OHGA, T., UCHIUMI, T., MAKINO, Y., KOIKE, K., WADA, M., KUWANO, M. & KOHNO, K. 1998. Direct involvement of the Y-box binding protein YB-1 in genotoxic stress-induced activation of the human multidrug resistance 1 gene. *J Biol Chem*, 273, 5997-6000.
- OHMORI, T., NAGAO, Y., MIZUKAMI, H., SAKATA, A., MURAMATSU, S. I., OZAWA, K., TOMINAGA, S. I., HANAZONO, Y., NISHIMURA, S., NUREKI, O. & SAKATA, Y. 2017. CRISPR/Cas9-mediated genome editing via postnatal administration of AAV vector cures haemophilia B mice. *Sci Rep*, 7, 4159.
- OKAMOTO, T., IZUMI, H., IMAMURA, T., TAKANO, H., ISE, T., UCHIUMI, T., KUWANO, M. & KOHNO, K. 2000. Direct interaction of p53 with the Y-box binding protein, YB-1: a mechanism for regulation of human gene expression. *Oncogene*, 19, 6194-202.
- OSBORN, M. J., WEBBER, B. R., KNIPPING, F., LONETREE, C. L., TENNIS, N., DEFEQ, A. P., MCELROY, A. N., STARKER, C. G., LEE, C., MERKEL, S., LUND, T. C., KELLY-SPRATT, K. S., JENSEN, M. C., VOYTAS, D. F., VON KALLE, C., SCHMIDT, M., GABRIEL, R., HIPPEL, K. L., MILLER, J. S., SCHARENBERG, A. M., TOLAR, J. & BLAZAR, B. R. 2016. Evaluation of TCR Gene Editing Achieved by TALENs, CRISPR/Cas9, and megaTAL Nucleases. *Mol Ther*, 24, 570-81.
- OSTERMAN, J. V., WADDELL, A. & APOSHIAN, H. V. 1970. DNA and gene therapy: uncoating of polyoma pseudovirus in mouse embryo cells. *Proc Natl Acad Sci U S A*, 67, 37-40.
- PADRON, E., BOWMAN, V., KALUDOV, N., GOVINDASAMY, L., LEVY, H., NICK, P., MCKENNA, R., MUZYCZKA, N., CHIORINI, J. A., BAKER, T. S. & AGBANDJE-MCKENNA, M. 2005. Structure of adeno-associated virus type 4. *J Virol*, 79, 5047-58.
- PAJUSOLA, K., GRUCHALA, M., JOCH, H., LUSCHER, T. F., YLA-HERTTUALA, S. & BUELER, H. 2002. Cell-type-specific characteristics modulate the transduction efficiency of adeno-associated virus type 2 and restrain infection of endothelial cells. *Journal of Virology*, 76, 11530-40.

- PARANJAPE, S. M. & HARRIS, E. 2007. Y box-binding protein-1 binds to the dengue virus 3'-untranslated region and mediates antiviral effects. *J Biol Chem*, 282, 30497-508.
- PARK, E. J., MIN, K. J., CHOI, K. S., KUBATKA, P., KRUZLIAK, P., KIM, D. E. & KWON, T. K. 2016. Chloroquine enhances TRAIL-mediated apoptosis through up-regulation of DR5 by stabilization of mRNA and protein in cancer cells. *Sci Rep*, 6, 22921.
- PARK, J. Y., LIM, B. P., LEE, K., KIM, Y. G. & JO, E. C. 2006. Scalable production of adeno-associated virus type 2 vectors via suspension transfection. *Biotechnol Bioeng*, 94, 416-30.
- PAVITT, G. D., RAMAIAH, K. V. A., KIMBALL, S. R. & HINNEBUSCH, A. G. 1998. eIF2 independently binds two distinct eIF2B subcomplexes that catalyze and regulate guanine-nucleotide exchange. *Genes Dev*, 12, 514-26.
- PEGORARO, G., MARCELLO, A., MYERS, M. P. & GIACCA, M. 2006. Regulation of adeno-associated virus DNA replication by the cellular TAF-I/set complex. *J Virol*, 80, 6855-64.
- PENAUD-BUDLOO, M., LE GUINER, C., NOWROUZI, A., TOROMANOFF, A., CHEREL, Y., CHENUAUD, P., SCHMIDT, M., VON KALLE, C., ROLLING, F., MOULLIER, P. & SNYDER, R. O. 2008. Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *J Virol*, 82, 7875-85.
- PEREZ-PINERA, P., KOCAK, D. D., VOCKLEY, C. M., ADLER, A. F., KABADI, A. M., POLSTEIN, L. R., THAKORE, P. I., GLASS, K. A., OUSTEROUT, D. G., LEONG, K. W., GUILAK, F., CRAWFORD, G. E., REDDY, T. E. & GERSBACH, C. A. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*, 10, 973-6.
- PIJLMAN, G. P., VAN SCHIJNDEL, J. E. & VLAK, J. M. 2003. Spontaneous excision of BAC vector sequences from bacmid-derived baculovirus expression vectors upon passage in insect cells. *J Gen Virol*, 84, 2669-78.
- PILDER, S., MOORE, M., LOGAN, J. & SHENK, T. 1986. The adenovirus E1B-55K transforming polypeptide modulates transport or cytoplasmic stabilization of viral and host cell mRNAs. *Mol Cell Biol*, 6, 470-6.
- PILLAI, D. R., LABBE, A. C., VANISAVETH, V., HONGVANGTHONG, B., POMPHIDA, S., INKATHONE, S., ZHONG, K. & KAIN, K. C. 2001. Plasmodium falciparum malaria in Laos: chloroquine treatment outcome and predictive value of molecular markers. *J Infect Dis*, 183, 789-95.
- PINGJUAN, L., P., K. B., Y., L. M., S., P. M., DANIEL, N.-G., H., G. S., A., P. E., KEITH, J. J. & QIN, L. 2018. Allele-Specific CRISPR-Cas9 Genome Editing of the Single-Base P23H Mutation for Rhodopsin-Associated Dominant Retinitis Pigmentosa. *The CRISPR Journal*, 1, 55-64.
- PIPAS, J. M. 1992. Common and unique features of T antigens encoded by the polyomavirus group. *J Virol*, 66, 3979-85.
- PISAREV, A. V., SKABKIN, M. A., THOMAS, A. A., MERRICK, W. C., OVCHINNIKOV, L. P. & SHATSKY, I. N. 2002. Positive and negative effects of the major mammalian messenger ribonucleoprotein p50 on binding of 40 S ribosomal subunits to the initiation codon of beta-globin mRNA. *J Biol Chem*, 277, 15445-51.
- POLLARD, H., REMY, J. S., LOUSSOUARN, G., DEMOLOMBE, S., BEHR, J. P. & ESCANDE, D. 1998. Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in mammalian cells. *J Biol Chem*, 273, 7507-11.
- POMBO, A., FERREIRA, J., BRIDGE, E. & CARMO-FONSECA, M. 1994. Adenovirus replication and transcription sites are spatially separated in the nucleus of infected cells. *Embo j*, 13, 5075-85.
- POPHAM, H. J., GASELA, J. J., GOODMAN, C. L. & MCINTOSH, A. H. 2010. Baculovirus infection influences host protein expression in two established insect cell lines. *J Insect Physiol*, 56, 1237-45.

- POTTER, M., CHESNUT, K., MUZYCZKA, N., FLOTTE, T. & ZOLOTUKHIN, S. 2002. Streamlined large-scale production of recombinant adeno-associated virus (rAAV) vectors. *Methods Enzymol*, 346, 413-30.
- POURCEL, C., SALVIGNOL, G. & VERGNAUD, G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151, 653-63.
- PRABHU, L., MUNDADE, R., WANG, B., WEI, H., HARTLEY, A.-V., MARTIN, M., MCELYEA, K., TEMM, C., SANDUSKY, G., LIU, Y. & LU, T. 2015. Critical role of phosphorylation of serine 165 of YBX1 on the activation of NF- κ B in colon cancer. *Oncotarget*, 6, 29396-412.
- PRASAD, K. M. & TREMPE, J. P. 1995. The adeno-associated virus Rep78 protein is covalently linked to viral DNA in a preformed virion. *Virology*, 214, 360-70.
- PUL, U., WURM, R., ARSLAN, Z., GEISSEN, R., HOFMANN, N. & WAGNER, R. 2010. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol*, 75, 1495-512.
- PUSHPARAJAN, C., CLAUS, J. D., MARSHALL, S. D. G. & VISNOVSKY, G. 2017. Nutritional demands and metabolic characteristics of the DSIR-HA-1179 insect cell line during growth and infection with the *Oryctes nudivirus*. *In Vitro Cell Dev Biol Anim*, 53, 908-21.
- QI, L. S., LARSON, M. H., GILBERT, L. A., DOUDNA, J. A., WEISSMAN, J. S., ARKIN, A. P. & LIM, W. A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152, 1173-83.
- QING, K., MAH, C., HANSEN, J., ZHOU, S., DWARKI, V. & SRIVASTAVA, A. 1999. Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat Med*, 5, 71-7.
- QIU, J. & BROWN, K. E. 1999. A 110-kDa nuclear shuttle protein, nucleolin, specifically binds to adeno-associated virus type 2 (AAV-2) capsid. *Virology*, 257, 373-82.
- QIU, J., NAYAK, R. & PINTEL, D. J. 2003. Alternative polyadenylation of adeno-associated virus type 5 RNA within an internal intron is governed by both a downstream element within the intron 3' splice acceptor and an element upstream of the P41 initiation site. *Journal of Virology*, 78, 83-93.
- QIU, P., SHANDILYA, H., D'ALESSIO, J. M., O'CONNOR, K., DUROCHER, J. & GERARD, G. F. 2004. Mutation detection using Surveyor nuclease. *Biotechniques*, 36, 702-7.
- RABINOWITZ, J. E., XIAO, W. & SAMULSKI, R. J. 1999. Insertional mutagenesis of AAV2 capsid and the production of recombinant virus. *Virology*, 265, 274-85.
- RAFFETSEDER, U., FRYE, B., RAUEN, T., JURCHOTT, K., ROYER, H. D., JANSEN, P. L. & MERTENS, P. R. 2003. Splicing factor SRp30c interaction with Y-box protein-1 confers nuclear YB-1 shuttling and alternative splice site selection. *J Biol Chem*, 278, 18241-8.
- RAHIM, A., URABE, M., MIZUKAMI, H., KUME, A., ICHIMARU, K. & OZAWA, K. 2011. Reduction of MBS85 gene expression after the targeted integration of a transgene into the AAVS1 site using adeno associated virus integration machinery. *International Journal of Genetics and Gene Therapy*, 1, 1-7.
- RALLAPALLI, P. M., KEMBALL-COOK, G., TUDDENHAM, E. G., GOMEZ, K. & PERKINS, S. J. 2013. An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *J Thromb Haemost*, 11, 1329-40.
- RAMASWAMY, S., TONNU, N., MENON, T., LEWIS, B. M., GREEN, K. T., WAMPLER, D., MONAHAN, P. E. & VERMA, I. M. 2018. Autologous and heterologous cell therapy for hemophilia B toward functional restoration of factor IX. *Cell Rep*, 23, 1565-80.

- RAN, F. A., CONG, L., YAN, W. X., SCOTT, D. A., GOOTENBERG, J. S., KRIZ, A. J., ZETSCHKE, B., SHALEM, O., WU, X., MAKAROVA, K. S., KOONIN, E. V., SHARP, P. A. & ZHANG, F. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 520, 186-91.
- RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154, 1380-9.
- RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013b. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 8, 2281-308.
- REDDING, S., STERNBERG, S. H., MARSHALL, M., GIBB, B., BHAT, P., GUEGLER, C. K., WIEDENHEFT, B., DOUDNA, J. A. & GREENE, E. C. 2015. Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell*, 163, 854-65.
- REED & MUENCH, H. 1938. A simple method of estimating fifty per cent endpoints. *Am. J. Epidemiol.*, 27, 493-7.
- REN, J., LIU, X., FANG, C., JIANG, S., JUNE, C. H. & ZHAO, Y. 2017. Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition. *Clin Cancer Res*, 23, 2255-66.
- RHIEL, M., MITCHELL-LOGEAN, C. & MURHAMMER, D. 1997. Comparison of *Trichoplusia ni* BTI-Tn-5B1-4 (high five) and *Spodoptera frugiperda* Sf-9 insect cell line metabolism in suspension cultures. *Biotechnol Bioeng*, 55, 909-20.
- RIBEIRO, S., MAIRHOFER, J., MADEIRA, C., DIOGO, M. M., SILVA, C. L. D., MONTEIRO, G., GRABHERR, R. & CABRAL, J. M. 2012. Plasmid DNA Size Does Affect Nonviral Gene Delivery Efficiency in Stem Cells. *Cellular Reprogramming*, 14, 130-137.
- RIDLEY, A. J. 2006. Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking. *Trends Cell Biol*, 16, 522-9.
- ROLLIE, C., SCHNEIDER, S., BRINKMANN, A. S., BOLT, E. L. & WHITE, M. F. 2015. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, 4.
- ROSENBLUTH, J. M., JOHNSON, K., TANG, L., TRIPLETT, T. & PIETENPOL, J. A. 2009. Evaluation of p63 and p73 antibodies for cross-reactivity. *Cell Cycle*, 8, 3702-6.
- ROUET, P., SMIH, F. & JASIN, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, 14, 8096-106.
- ROUILLON, C., ZHOU, M., ZHANG, J., POLITIS, A., BEILSTEN-EDMANDS, V., CANNONE, G., GRAHAM, S., ROBINSON, C. V., SPAGNOLO, L. & WHITE, M. F. 2013. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell*, 52, 124-34.
- RUDIN, N., SUGARMAN, E. & HABER, J. E. 1989. Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics*, 122, 519-34.
- RUFFING, M., ZENTGRAF, H. & KLEINSCHMIDT, J. A. 1992. Assembly of viruslike particles by recombinant structural proteins of adeno-associated virus type 2 in insect cells. *J Virol*, 66, 6922-30.
- RUPP, L. J., SCHUMANN, K., ROYBAL, K. T., GATE, R. E., YE, C. J., LIM, W. A. & MARSON, A. 2017. CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci Rep*, 7, 737.
- RUTLEDGE, E. A., HALBERT, C. L. & RUSSELL, D. W. 1998. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J Virol*, 72, 309-19.
- SAFAK, M., GALLIA, G. L., ANSARI, S. A. & KHALILI, K. 1999. Physical and functional interaction between the Y-box binding protein YB-1 and human polyomavirus JC virus large T antigen. *J Virol*, 73, 10146-57.
- SALEH-GOHARI, N. & HELLEDAY, T. 2004. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res*, 32, 3683-8.

- SAMPSON, T. R. & WEISS, D. S. 2013. Cas9-dependent endogenous gene regulation is required for bacterial virulence. *Biochem Soc Trans*, 41, 1407-11.
- SAMULSKI, R. J., BERNS, K. I., TAN, M. & MUZYCZKA, N. 1982. Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc Natl Acad Sci U S A*, 79, 2077-81.
- SAMULSKI, R. J., CHANG, L. S. & SHENK, T. 1987. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. *J Virol*, 61, 3096-101.
- SAMULSKI, R. J. & SHENK, T. 1988. Adenovirus E1B 55-Mr polypeptide facilitates timely cytoplasmic accumulation of adeno-associated virus mRNAs. *J Virol*, 62, 206-10.
- SAMULSKI, R. J., ZHU, X., XIAO, X., BROOK, J. D., HOUSMAN, D. E., EPSTEIN, N. & HUNTER, L. A. 1991. Targeted integration of Adeno-associated virus (AAV) into human chromosome 19. *EMBO J*, 10, 3941-50.
- SANDS, M. S. 2011. AAV-mediated liver-directed gene therapy. *Methods Mol Biol*, 807, 141-57.
- SANJANA, N. E., SHALEM, O. & ZHANG, F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods*, 11, 783-4.
- SANLIOGLU, S., BENSON, P. K., YANG, J., ATKINSON, E. M., REYNOLDS, T. & ENGELHARDT, J. F. 2000. Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by rac1 and phosphatidylinositol-3 kinase activation. *J Virol*, 74, 9184-96.
- SATKUNANATHAN, S., THORPE, R. & ZHAO, Y. 2017. The function of DNA binding protein nucleophosmin in AAV replication. *Virology*, 510, 46-54.
- SATKUNANATHAN, S., WHEELER, J., THORPE, R. & ZHAO, Y. 2014. Establishment of a novel cell line for the enhanced production of recombinant adeno-associated virus vectors for gene therapy. *Hum Gene Ther*, 25, 929-41.
- SAUPE, M., RAUSCHENBERGER, L., PREUSS, M., OSWALD, S., FUSSEK, S., ZIMMERMANN, U., WALTHER, R., KNABBE, C., BURCHARDT, M. & STOPE, M. B. 2015. Differential expression of the multidrug resistance 1 (MDR1) protein in prostate cancer cells is independent from anticancer drug treatment and Y box binding protein 1 (YB-1) activity. *World J Urol*, 33, 1481-6.
- SAWAYA, B. E., KHALILI, K. & AMINI, S. 1998. Transcription of the human immunodeficiency virus type 1 (HIV-1) promoter in central nervous system cells: effect of YB-1 on expression of the HIV-1 long terminal repeat. *J Gen Virol*, 79, 239-46.
- SCHERER, P. E., LEDERKREMER, G. Z., WILLIAMS, S., FOGLIANO, M., BALDINI, G. & LODISH, H. F. 1996. Cab45, a novel (Ca²⁺)-binding protein localized to the Golgi lumen. *J Cell Biol*, 133, 257-68.
- SCHIEDNER, G., HERTEL, S., BIALEK, C., KEWES, H., WASCHUTZA, G. & VOLPERS, C. 2008. Efficient and reproducible generation of high-expressing, stable human cell lines without need for antibiotic selection. *BMC Biotechnol*, 8, 13.
- SCHITTEK, B., PSENNER, K., SAUER, B., MEIER, F., IFTNER, T. & GARBE, C. 2007. The increased expression of Y box-binding protein 1 in melanoma stimulates proliferation and tumor invasion, antagonizes apoptosis and enhances chemoresistance. *Int J Cancer*, 120, 2110-8.
- SCHMIDT, M., AFIONE, S. & KOTIN, R. M. 2000. Adeno-associated virus type 2 Rep78 induces apoptosis through caspase activation independently of p53. *J Virol*, 74, 9441-50.
- SCHMIDT, M., GROT, E., CERVENKA, P., WAINER, S., BUCK, C. & CHIORINI, J. A. 2006. Identification and characterization of novel adeno-associated virus isolates in ATCC virus stocks. *J Virol*, 80, 5082-5.
- SCHMUKLER, E., WOLFSON, E., HAKLAI, R., ELAD-SFADIA, G., KLOOG, Y. & PINKAS-KRAMARSKI, R. 2014. Chloroquine synergizes with FTS to enhance cell growth inhibition and cell death. *Oncotarget*, 5, 173-84.

- SCHREIBER, C. A., SAKUMA, T., IZUMIYA, Y., HOLDITCH, S. J., HICKEY, R. D., BRESSIN, R. K., BASU, U., KOIDE, K., ASOKAN, A. & IKEDA, Y. 2015. An siRNA Screen Identifies the U2 snRNP Spliceosome as a Host Restriction Factor for Recombinant Adeno-associated Viruses. *PLoS Pathog*, 11, e1005082.
- SCHWARTZ, R. A., LAKDAWALA, S. S., ESHLEMAN, H. D., RUSSELL, M. R., CARSON, C. T. & WEITZMAN, M. D. 2008. Distinct requirements of adenovirus E1b55K protein for degradation of cellular substrates. *J Virol*, 82, 9043-55.
- SEEGER, C. & SOHN, J. A. 2014. Targeting Hepatitis B virus with CRISPR/Cas9. *Mol Ther Nucleic Acids*, 3, e216.
- SEISENBERGER, G., RIED, M. U., ENDRESS, T., BUNING, H., HALLEK, M. & BRAUCHLE, C. 2001. Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science*, 294, 1929-32.
- SEKI, A. & RUTZ, S. 2018. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *J Exp Med*, 215, 985-97.
- SHEN, B., ZHANG, W., ZHANG, J., ZHOU, J., WANG, J., CHEN, L., WANG, L., HODGKINS, A., IYER, V., HUANG, X. & SKARNES, W. C. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*, 11, 399-402.
- SHI, J. H., CUI, N. P., WANG, S., ZHAO, M. Z., WANG, B., WANG, Y. N. & CHEN, B. P. 2016. Overexpression of YB1 C-terminal domain inhibits proliferation, angiogenesis and tumorigenicity in a SK-BR-3 breast cancer xenograft mouse model. *FEBS Open Bio*, 6, 33-42.
- SHI, J. H., ZHENG, B., CHEN, S., MA, G. Y. & WEN, J. K. 2012. Retinoic acid receptor alpha mediates all-trans-retinoic acid-induced Klf4 gene expression by regulating Klf4 promoter activity in vascular smooth muscle cells. *J Biol Chem*, 287, 10799-811.
- SHI, J. H., ZHENG, B., LI, Y. H., SUN, Y., HAN, A. L., ZHANG, X. H., LV, X. R., CHEN, S. & WEN, J. K. 2013. Novel insight into Y-box binding protein 1 in the regulation of vascular smooth muscle cell proliferation through targeting GC box-dependent genes. *FEBS Lett*, 587, 1326-32.
- SHIBAHARA, K., UCHIUMI, T., FUKUDA, T., KURA, S., TOMINAGA, Y., MAEHARA, Y., KOHNO, K., NAKABEPPU, Y., TSUZUKI, T. & KUWANO, M. 2004. Targeted disruption of one allele of the Y-box binding protein-1 (YB-1) gene in mouse embryonic stem cells and increased sensitivity to cisplatin and mitomycin C. *Cancer Sci*, 95, 348-53.
- SHIN, H. Y., WANG, C., LEE, H. K., YOO, K. H., ZENG, X., KUHNS, T., YANG, C. M., MOHR, T., LIU, C. & HENNIGHAUSEN, L. 2017. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun*, 8, 15464.
- SHIOTA, M., IZUMI, H., ONITSUKA, T., MIYAMOTO, N., KASHIWAGI, E., KIDANI, A., YOKOMIZO, A., NAITO, S. & KOHNO, K. 2008. Twist promotes tumor cell growth through YB-1 expression. *Cancer Res*, 68, 98-105.
- SHIRAI, A., MATSUYAMA, A., YASHIRODA, Y., HASHIMOTO, A., KAWAMURA, Y., ARAI, R., KOMATSU, Y., HORINOUCI, S. & YOSHIDA, M. 2008. Global analysis of gel mobility of proteins and its use in target identification. *J Biol Chem*, 283, 10745-52.
- SHIRYAEV, S. A., MESCI, P., PINTO, A., FERNANDES, I., SHEETS, N., SHRESTA, S., FARHY, C., HUANG, C. T., STRONGIN, A. Y., MUOTRI, A. R. & TERSKIKH, A. V. 2017. Repurposing of the anti-malaria drug chloroquine for Zika Virus treatment and prophylaxis. *Sci Rep*, 7, 15771.
- SHMAKOV, S. A., SITNIK, V., MAKAROVA, K. S., WOLF, Y. I., SEVERINOV, K. V. & KOONIN, E. V. 2017. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio*, 8.
- SIJEN, T., FLEENOR, J., SIMMER, F., THIJSEN, K. L., PARRISH, S., TIMMONS, L., PLASTERK, R. H. & FIRE, A. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, 107, 465-76.

- SILVA, W., CARDOSO, C., RIBEIRO, A. F., TERRA, W. R. & FERREIRA, C. 2013. Midgut proteins released by microapocrine secretion in *Spodoptera frugiperda*. *J Insect Physiol*, 59, 70-80.
- SIMOSSIS, V. A., KLEINJUNG, J. & HERINGA, J. 2005. Homology-extended sequence alignment. *Nucleic Acids Res*, 33, 816-24.
- SINNBERG, T., SAUER, B., HOLM, P., SPANGLER, B., KUPHAL, S., BOSSERHOFF, A. & SCHITTEK, B. 2012. MAPK and PI3K/AKT mediated YB-1 activation promotes melanoma cell proliferation which is counteracted by an autoregulatory loop. *Exp Dermatol*, 21, 265-70.
- SKABKIN, M. A., EVDOKIMOVA, V., THOMAS, A. A. & OVCHINNIKOV, L. P. 2001. The major messenger ribonucleoprotein particle protein p50 (YB-1) promotes nucleic acid strand annealing. *J Biol Chem*, 276, 44841-7.
- SKABKIN, M. A., KISELYOVA, O. I., CHERNOV, K. G., SOROKIN, A. V., DUBROVIN, E. V., YAMINSKY, I. V., VASILIEV, V. D. & OVCHINNIKOV, L. P. 2004. Structural organization of mRNA complexes with major core mRNP protein YB-1. *Nucleic Acids Res*, 32, 5621-35.
- SLAYMAKER, I. M., GAO, L., ZETSCHKE, B., SCOTT, D. A., YAN, W. X. & ZHANG, F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351, 84-8.
- SMITH, G. E., FRASER, M. J. & SUMMERS, M. D. 1983a. Molecular engineering of the *Autographa californica* nuclear polyhedrosis virus genome: deletion mutations within the polyhedrin gene. *J Virol*, 46, 584-93.
- SMITH, G. E., SUMMERS, M. D. & FRASER, M. J. 1983b. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol*, 3, 2156-65.
- SMITH, R. H. & KOTIN, R. M. 2000. An adeno-associated Virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. *Journal of Virology*, 74, 3122-9.
- SMITH, R. H., LEVY, J. R. & KOTIN, R. M. 2009. A simplified baculovirus-AAV expression vector system coupled with one-step affinity purification yields high-titer rAAV stocks from insect cells. *Mol Ther*, 17, 1888-96.
- SOLOMON, D. A., KIM, T., DIAZ-MARTINEZ, L. A., FAIR, J., ELKAHLOUN, A. G., HARRIS, B. T., TORETSKY, J. A., ROSENBERG, S. A., SHUKLA, N., LADANYI, M., SAMUELS, Y., JAMES, C. D., YU, H., KIM, J. S. & WALDMAN, T. 2011. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*, 333, 1039-43.
- SOMASEKHARAN, S. P., EL-NAGGAR, A., LEPRIVIER, G., CHENG, H., HAJEE, S., GRUNEWALD, T. G., ZHANG, F., NG, T., DELATTRE, O., EVDOKIMOVA, V., WANG, Y., GLEAVE, M. & SORENSEN, P. H. 2015. YB-1 regulates stress granule formation and tumor progression by translationally activating G3BP1. *J Cell Biol*, 208, 913-29.
- SONAWANE, N. D., SZOKA, F. C., JR. & VERKMAN, A. S. 2003. Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. *J Biol Chem*, 278, 44826-31.
- SONNTAG, F., KOTHER, K., SCHMIDT, K., WEGHOFER, M., RAUPP, C., NIETO, K., KUCK, A., GERLACH, B., BOTTCHER, B., MULLER, O. J., LUX, K., HORER, M. & KLEINSCHMIDT, J. A. 2011. The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. *J Virol*, 85, 12686-97.
- SONNTAG, F., SCHMIDT, K. & KLEINSCHMIDT, J. A. 2010. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc Natl Acad Sci U S A*, 107, 10220-5.
- SORENSEN, K. S., MAHANEY, B. L., LEES-MILLER, S. P. & COBB, J. A. 2017. The non-homologous end-joining factor Nej1 inhibits resection mediated by Dna2-Sgs1 nuclease-helicase at DNA double strand breaks. *J Biol Chem*, 292, 14576-86.
- SOROKIN, A. V., SELYUTINA, A. A., SKABKIN, M. A., GURYANOV, S. G., NAZIMOV, I. V., RICHARD, C., TH'NG, J., YAU, J., SORENSEN, P. H., OVCHINNIKOV, L. P. & EVDOKIMOVA, V. 2005.

- Proteasome-mediated cleavage of the Y-box-binding protein 1 is linked to DNA-damage stress response. *Embo j*, 24, 3602-12.
- SPITKOVSKY, D., JANSEN-DURR, P., KARSENTI, E. & HOFFMAN, I. 1996. S-phase induction by adenovirus E1A requires activation of cdc25a tyrosine phosphatase. *Oncogene*, 12, 2549-54.
- SRIVASTAVA, A., LUSBY, E. W. & BERNS, K. I. 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol*, 45, 555-64.
- STAALS, R. H. J., AGARI, Y., MAKI-YONEKURA, S., ZHU, Y., TAYLOR, D. W., VAN DUIJN, E., BARENDREGT, A., VLOT, M., KOEHORST, J. J., SAKAMOTO, K., MASUDA, A., DOHMAE, N., SCHAAP, P. J., DOUDNA, J. A., HECK, A. J. R., YONEKURA, K., VAN DER OOST, J. & SHINKAI, A. 2013. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell*, 52, 135-45.
- STAHNKE, S., LUX, K., UHRIG, S., KREPEL, F., HOSEL, M., COUTELLE, O., OGRIS, M., HALLEK, M. & BUNING, H. 2011. Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles. *Virology*, 409, 77-83.
- STAIB, C., PESCH, J., GERWIG, R., GERBER, J. K., BREHM, U., STANGL, A. & GRUMMT, F. 1996. p53 inhibits JC virus DNA replication in vivo and interacts with JC virus large T-antigen. *Virology*, 219, 237-46.
- STEELE, K. H., STONE, B. J., FRANKLIN, K. M., FATH-GOODIN, A., ZHANG, X., JIANG, H., WEBB, B. A. & GEISLER, C. 2017. Improving the baculovirus expression vector system with vankyrin-enhanced technology. *Biotechnol Prog*, 33, 1496-507.
- STEINBACH, S., WISTUBA, A., BOCK, T. & KLEINSCHMIDT, J. A. 1997. Assembly of adeno-associated virus type 2 capsids in vitro. *J Gen Virol*, 78, 1453-62.
- STERN, A., KEREN, L., WURTZEL, O., AMITAI, G. & SOREK, R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet*, 26, 335-40.
- STICKELER, E., FRASER, S. D., HONIG, A., CHEN, A. L., BERGET, S. M. & COOPER, T. A. 2001. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *The EMBO Journal*, 20, 3821-30.
- STONE, R., HAYASHI, T., BAJIMAYA, S., HODGES, E. & TAKIMOTO, T. 2016. Critical role of Rab11a-mediated recycling endosomes in the assembly of type I parainfluenza viruses. *Virology*, 487, 11-8.
- STRACKER, T. H., CASSELL, G. D., WARD, P., LOO, Y. M., VAN BREUKELLEN, B., CARRINGTON-LAWRENCE, S. D., HAMATAKE, R. K., VAN DER VLIET, P. C., WELLER, S. K., MELENDY, T. & WEITZMAN, M. D. 2003. The Rep Protein of Adeno-Associated Virus Type 2 Interacts with Single-Stranded DNA-Binding Proteins That Enhance Viral Replication. *Journal of Virology*, 78, 441-453.
- STRATFORD, A., J FRY, C., DESILETS, C., H DAVIES, A., CHO, Y., LI, Y., DONG, Z., BERQUIN, I., ROUX, P. & DUNN, S. 2008. Y-box binding protein-1 serine 102 is a downstream target of p90 ribosomal S6 kinase in basal-like breast cancer cells. *Breast Cancer Res*, 10, R99.
- STRAUS, S. E., SEBRING, E. D. & ROSE, J. A. 1976. Concatemers of alternating plus and minus strands are intermediates in adenovirus-associated virus DNA synthesis. *Proc Natl Acad Sci U S A*, 73, 742-6.
- STRAYER, D. S. 1996. SV40 as an effective gene transfer vector *in vivo*. *J Biol Chem*, 271, 24741-6.
- STROBEL, B., MILLER, F. D., RIST, W. & LAMLA, T. 2015. Comparative Analysis of Cesium Chloride- and Iodixanol-Based Purification of Recombinant Adeno-Associated Viral Vectors for Preclinical Applications. *Hum Gene Ther Methods*, 26, 147-57.
- SU, S., HU, B., SHAO, J., SHEN, B., DU, J., DU, Y., ZHOU, J., YU, L., ZHANG, L., CHEN, F., SHA, H., CHENG, L., MENG, F., ZOU, Z., HUANG, X. & LIU, B. 2016a. CRISPR-Cas9 mediated efficient PD-1 disruption on human primary T cells from cancer patients. *Sci Rep*, 6, 20070.

- SU, T., LIU, F., GU, P., JIN, H., CHANG, Y., WANG, Q., LIANG, Q. & QI, Q. 2016b. A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. *Sci Rep*, 6, 37895.
- SUMMERFORD, C. & SAMULSKI, R. J. 1998. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol*, 72, 1438-45.
- SUN, D., SAHU, B., GAO, S., SCHUR, R. M., VAIDYA, A. M., MAEDA, A., PALCZEWSKI, K. & LU, Z.-R. 2017. Targeted Multifunctional Lipid ECO Plasmid DNA Nanoparticles as Efficient Non-viral Gene Therapy for Leber's Congenital Amaurosis. *Molecular therapy. Nucleic acids*, 7, 42-52.
- SUROSKY, R. T., URABE, M., GODWIN, S. G., MCQUISTON, S. A., KURTZMAN, G. J., OZAWA, K. & NATSOULIS, G. 1997. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J Virol*, 71, 7951-9.
- SWAMYNATHAN, S. K., VARMA, B. R., WEBER, K. T. & GUNTAKA, R. V. 2002. Targeted disruption of one allele of the Y-box protein gene, Chk-YB-1b, in DT40 cells results in major defects in cell cycle. *Biochem Biophys Res Commun*, 296, 451-7.
- SZYBALSKA, E. H. & SZYBALSKI, S. W. 1962. Genetics of human cell lines, IV. DNA-mediated heritable transformation of a biochemical trait. *Proc Natl Acad Sci U S A*, 48, 2026-34.
- TABEBORDBAR, M., ZHU, K., CHENG, J. K. W., CHEW, W. L., WIDRICK, J. J., YAN, W. X., MAESNER, C., WU, E. Y., XIAO, R., RAN, F. A., CONG, L., ZHANG, F., VANDENBERGHE, L. H., CHURCH, G. M. & WAGERS, A. J. 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science*, 351, 407-11.
- TAKIYA, S., NISHITA, Y., ISHIKAWA, S., OHNO, K., TAMURA, T.-A. & SUZUKI, Y. 2004. Bombyx Y-Box protein BYB facilitates specific DNA interaction of various DNA binding proteins independently of the cold shock domain. *J Biochem*, 135, 683-93.
- TANABE, Y., NAGATOISHI, S. & TSUMOTO, K. 2015. Thermodynamic characterization of the interaction between the human Y-box binding protein YB-1 and nucleic acids. *Mol Biosyst*, 11, 2441-8.
- TAY, M., LIU, S. & YUAN, Y. A. 2015. Crystal structure of Thermobifida fusca Cse1 reveals target DNA binding site. *Protein Sci*, 24, 236-45.
- TEBAS, P., STEIN, D., TANG, W. W., FRANK, I., WANG, S. Q., LEE, G., SPRATT, S. K., SUROSKY, R. T., GIEDLIN, M. A., NICHOL, G., HOLMES, M. C., GREGORY, P. D., ANDO, D. G., KALOS, M., COLLMAN, R. G., BINDER-SCHOLL, G., PLESA, G., HWANG, W. T., LEVINE, B. L. & JUNE, C. H. 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*, 370, 901-10.
- TERAO, Y., MORI, Y., YAMAGUCHI, M., SHIMIZU, Y., OOE, K., HAMADA, S. & KAWABATA, S. 2008. Group A streptococcal cysteine protease degrades C3 (C3b) and contributes to evasion of innate immunity. *J Biol Chem*, 283, 6253-60.
- TRATSCHIN, J. D., WEST, M. H., SANDBANK, T. & CARTER, B. J. 1984. A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase. *Mol Cell Biol*, 4, 2072-81.
- TRILLING, D. M. & AXELROD, D. 1970. Encapsidation of free host DNA by simian virus 40: a simian virus 40 pseudovirus. *Science*, 168, 268-71.
- TROCK, B. J., LEONESSA, F. & CLARKE, R. 1997. Multidrug resistance in breast cancer: a meta-analysis of MDR1/gp170 expression and its possible functional significance. *J Natl Cancer Inst*, 89, 917-31.
- TSAI, S. Q., ZHENG, Z., NGUYEN, N. T., LIEBERS, M., TOPKAR, V. V., THAPAR, V., WYVEKENS, N., KHAYTER, C., IAFRATE, A. J., LE, L. P., ARYEE, M. J. & JOUNG, J. K. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, 33, 187-97.

- TSAO, E. I., MASON, M. R., CACCIUTTOLO, M. A., BOWEN, S. H. & FOLENA-WASSERMAN, G. 1996. Production of parvovirus B19 vaccine in insect cells co-infected with double baculoviruses. *Biotechnol Bioeng*, 49, 130-8.
- TSE, L. V., KLINC, K. A., MADIGAN, V. J., CASTELLANOS RIVERA, R. M., WELLS, L. F., HAVLIK, L. P., SMITH, J. K., AGBANDJE-MCKENNA, M. & ASOKAN, A. 2017. Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proc Natl Acad Sci U S A*, 114, E4812-21.
- UCHIUMI, T., FOTOVATI, A., SASAGURI, T., SHIBAHARA, K., SHIMADA, T., FUKUDA, T., NAKAMURA, T., IZUMI, H., TSUZUKI, T., KUWANO, M. & KOHNO, K. 2006. YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation. *J Biol Chem*, 281, 40440-9.
- URABE, M., DING, C. & KOTIN, R. M. 2002. Insect cells as a factory to produce adeno-associated virus type 2 vectors. *Hum Gene Ther*, 13, 1935-43.
- URABE, M., NAKAKURA, T., XIN, K. Q., OBARA, Y., MIZUKAMI, H., KUME, A., KOTIN, R. M. & OZAWA, K. 2006. Scalable generation of high-titer recombinant adeno-associated virus type 5 in insect cells. *J Virol*, 80, 1874-85.
- URNOV, F. D., MILLER, J. C., LEE, Y. L., BEAUSEJOUR, C. M., ROCK, J. M., AUGUSTUS, S., JAMIESON, A. C., PORTEUS, M. H., GREGORY, P. D. & HOLMES, M. C. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 425, 646-51.
- USAMI, A., ISHIYAMA, S., ENOMOTO, C., OKAZAKI, H., HIGUCHI, K., IKEDA, M., YAMAMOTO, T., SUGAI, M., ISHIKAWA, Y., HOSAKA, Y., KOYAMA, T., TOBITA, Y., EBIHARA, S., MOCHIZUKI, T., ASANO, Y. & NAGAYA, H. 2011. Comparison of recombinant protein expression in a baculovirus system in insect cells (Sf9) and silkworm. *J Biochem*, 149, 219-27.
- UTSUNO, K. & ULUDAG, H. 2010. Thermodynamics of polyethylenimine-DNA binding and DNA condensation. *Biophys J*, 99, 201-7.
- VAN AELST, L. & D'SOUZA-SCHOREY, C. 1997. Rho GTPases and signaling networks. *Genes Dev*, 11, 2295-322.
- VAN DEN BOSSCHE, A., HARDWICK, S. W., CEYSSENS, P. J., HENDRIX, H., VOET, M., DENDOOVEN, T., BANDYRA, K. J., DE MAEYER, M., AERTSEN, A., NOBEN, J. P., LUISI, B. F. & LAVIGNE, R. 2016. Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome. *Elife*, 5.
- VAN DIEMEN, F. R., KRUSE, E. M., HOOYKAAS, M. J., BRUGGELING, C. E., SCHURCH, A. C., VAN HAM, P. M., IMHOF, S. M., NIJHUIS, M., WIERTZ, E. J. & LEBBINK, R. J. 2016. CRISPR/Cas9-Mediated Genome Editing of Herpesviruses Limits Productive and Latent Infections. *PLoS Pathog*, 12, e1005701.
- VAN ERP, P. B., JACKSON, R. N., CARTER, J., GOLDEN, S. M., BAILEY, S. & WIEDENHEFT, B. 2015. Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic Acids Res*, 43, 8381-91.
- VAN GESTEL, M. A., BOENDER, A. J., DE VRIND, V. A., GARNER, K. M., LUIJENDIJK, M. C. & ADAN, R. A. 2014. Recombinant adeno-associated virus: efficient transduction of the rat VMH and clearance from blood. *PLoS One*, 9, e97639.
- VAN LOENEN, M. M., DE BOER, R., AMIR, A. L., HAGEDOORN, R. S., VOLBEDA, G. L., WILLEMZE, R., VAN ROOD, J. J., FALKENBURG, J. H. & HEEMSKERK, M. H. 2010. Mixed T cell receptor dimers harbor potentially harmful neoreactivity. *Proc Natl Acad Sci U S A*, 107, 10972-7.
- VAN ROEYEN, C. R., EITNER, F., MARTINKUS, S., THIELTGES, S. R., OSTENDORF, T., BOKEMEYER, D., LUSCHER, B., LUSCHER-FIRZLAFF, J. M., FLOEGE, J. & MERTENS, P. R. 2005. Y-box protein 1 mediates PDGF-B effects in mesangioproliferative glomerular disease. *J Am Soc Nephrol*, 16, 2985-96.

- VAN ROEYEN, C. R., SCURT, F. G., BRANDT, S., KUHL, V. A., MARTINKUS, S., DJUDJAJ, S., RAFFETSEDER, U., ROYER, H. D., STEFANIDIS, I., DUNN, S. E., DOOLEY, S., WENG, H., FISCHER, T., LINDQUIST, J. A. & MERTENS, P. R. 2013. Cold shock Y-box protein-1 proteolysis autoregulates its transcriptional activities. *Cell Commun Signal*, 11, 63.
- VAN VALEN, L. 1973. *A New Evolutionary Law*.
- VANDENBERGHE, L. H., XIAO, R., LOCK, M., LIN, J., KORN, M. & WILSON, J. M. 2010. Efficient serotype-dependent release of functional vector into the culture medium during adeno-associated virus manufacturing. *Hum Gene Ther*, 21, 1251-7.
- VAUGHN, J. L., GOODWIN, R. H., TOMPKINS, G. J. & MCCAWLEY, P. 1977. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro*, 13, 213-7.
- VENKATAKRISHNAN, B., YARBROUGH, J., DOMSIC, J., BENNETT, A., BOTHNER, B., KOZYREVA, O. G., SAMULSKI, R. J., MUZYCZKA, N., MCKENNA, R. & AGBANDJE-MCKENNA, M. 2013. Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. *J Virol*, 87, 4974-84.
- VERA, M., PRIETO, J., STRAYER, D. S. & FORTES, P. 2004. Factors influencing the production of recombinant SV40 vectors. *Mol Ther*, 10, 780-91.
- VERA, M., SOBREVALS, L., ZARATIEGUI, M., MARTINEZ, L., PALENCIA, B., RODRIGUEZ, C. M., PRIETO, J. & FORTES, P. 2007. Liver transduction with a simian virus 40 vector encoding insulin-like growth factor I reduces hepatic damage and the development of liver cirrhosis. *Gene Ther*, 14, 203-10.
- VERON, P., LEBORGNE, C., MONTEILHET, V., BOUTIN, S., MARTIN, S., MOULLIER, P. & MASURIER, C. 2012. Humoral and cellular capsid-specific immune responses to adeno-associated virus type 1 in randomized healthy donors. *J Immunol*, 188, 6418-24.
- VERSCHOOTEN, L., BARRETTE, K., VAN KELST, S., RUBIO ROMERO, N., PROBY, C., DE VOS, R., AGOSTINIS, P. & GARMYN, M. 2012. Autophagy inhibitor chloroquine enhanced the cell death inducing effect of the flavonoid luteolin in metastatic squamous cell carcinoma cells. *PLoS One*, 7, e48264.
- VESSONI, A. T., QUINET, A., DE ANDRADE-LIMA, L. C., MARTINS, D. J., GARCIA, C. C., ROCHA, C. R., VIEIRA, D. B. & MENCK, C. F. 2016. Chloroquine-induced glioma cells death is associated with mitochondrial membrane potential loss, but not oxidative stress. *Free Radic Biol Med*, 90, 91-100.
- VOLPERS, C. & KOCHANNEK, S. 2004. Adenoviral vectors for gene transfer and therapy. *J Gene Med*, 6 Suppl 1, S164-71.
- VOUILLOT, L., THÉLIE, A. & POLLET, N. 2015. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3 (Bethesda)*, 5, 407-15.
- VRLJIC, M., STROP, P., HILL, R. C., HANSEN, K. C., CHU, S. & BRUNGER, A. T. 2011. Post-translational modifications and lipid binding profile of insect cell-expressed full-length mammalian synaptotagmin 1. *Biochemistry*, 50, 9998-10012.
- WAKEFIELD, N., RAJAN, R. & SONTHEIMER, E. J. 2015. Primary processing of CRISPR RNA by the endonuclease Cas6 in *Staphylococcus epidermidis*. *FEBS Lett*, 589, 3197-204.
- WALTERS, R. W., AGBANDJE-MCKENNA, M., BOWMAN, V. D., MONINGER, T. O., OLSON, N. H., SEILER, M., CHIORINI, J. A., BAKER, T. S. & ZABNER, J. 2004. Structure of adeno-associated virus serotype 5. *Journal of Virology*, 78, 3361-71.
- WANG, D., MOU, H., LI, S., LI, Y., HOUGH, S., TRAN, K., LI, J., YIN, H., ANDERSON, D. G., SONTHEIMER, E. J., WENG, Z., GAO, G. & XUE, W. 2015a. Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum Gene Ther*, 26, 432-42.
- WANG, G., ZHAO, N., BERKHOUT, B. & DAS, A. T. 2016a. CRISPR-Cas9 Can Inhibit HIV-1 Replication but NHEJ Repair Facilitates Virus Escape. *Mol Ther*, 24, 522-6.

- WANG, H., SUN, R., GU, M., LI, S., ZHANG, B., CHI, Z. & HAO, L. 2015b. shRNA-Mediated Silencing of Y-Box Binding Protein-1 (YB-1) Suppresses Growth of Neuroblastoma Cell SH-SY5Y In Vitro and In Vivo. *PLoS One*, 10, e0127224.
- WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153, 910-8.
- WANG, J., LI, J., ZHAO, H., SHENG, G., WANG, M., YIN, M. & WANG, Y. 2015c. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell*, 163, 840-53.
- WANG, J. & QUAKE, S. R. 2014. RNA-guided endonuclease provides a therapeutic strategy to cure latent herpesviridae infection. *Proc Natl Acad Sci U S A*, 111, 13157-62.
- WANG, J. Y., DEL VALLE, L., PERUZZI, F., TROJANEK, J., GIORDANO, A., KHALILI, K. & REISS, K. 2004a. Polyomaviruses and cancer--interplay between viral proteins and signal transduction pathways. *J Exp Clin Cancer Res*, 23, 373-83.
- WANG, K., HUANG, S., KAPOOR-MUNSHI, A. & NEMEROW, G. 1998. Adenovirus internalization and infection require dynamin. *J Virol*, 72, 3455-8.
- WANG, L., NICHOLS, T. C., READ, M. S., BELLINGER, D. A. & VERMA, I. M. 2000a. Sustained expression of therapeutic level of factor IX in hemophilia B dogs by AAV-mediated gene therapy in liver. *Mol Ther*, 1, 154-8.
- WANG, N., YAMANAKA, K. & INOUE, M. 2000b. Acquisition of double-stranded DNA-binding ability in a hybrid protein between Escherichia coli CspA and the cold shock domain of human YB-1. *Molecular Microbiology*, 38, 526-34.
- WANG, R., LI, M., GONG, L., HU, S. & XIANG, H. 2016b. DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res*, 44, 4266-77.
- WANG, W., JIA, Y. L., LI, Y. C., JING, C. Q., GUO, X., SHANG, X. F., ZHAO, C. P. & WANG, T. Y. 2017. Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells. *Sci Rep*, 7, 10416.
- WANG, W., WANG, H. J., WANG, B., LI, Y., QIN, Y., ZHENG, L. S., ZHOU, J. S., QU, P. H., SHI, J. H. & ZHANG, H. S. 2016c. The role of the Y Box binding protein 1 C-terminal domain in vascular endothelial cell proliferation, apoptosis, and angiogenesis. *DNA Cell Biol*, 35, 24-32.
- WANG, X. S., PONNAZHAGAN, S. & SRIVASTAVA, A. 1996. Rescue and replication of adeno-associated virus type 2 as well as vector DNA sequences from recombinant plasmids containing deletions in the viral inverted terminal repeats: selective encapsidation of viral genomes in progeny virions. *J Virol*, 70, 1668-77.
- WANG, Y., ZHANG, H., GAO, Y., HUANG, C., ZHOU, A., ZHOU, Y. & LI, Y. 2015d. Sequential post-translational modifications regulate PKC degradation. 27, 410-20.
- WANG, Z., HALBERT, C. L., LEE, D., BUTTS, T., TAPSCOTT, S. J., STORB, R. & MILLER, A. D. 2014. Elimination of contaminating cap genes in AAV vector virions reduces immune responses and improves transgene expression in a canine gene therapy model. *Gene Ther*, 21, 363-70.
- WANG, Z., PAN, Q., GENDRON, P., ZHU, W., GUO, F., CEN, S., WAINBERG, M. A. & LIANG, C. 2016d. CRISPR/Cas9-derived mutations both Inhibit HIV-1 replication and accelerate viral escape. *Cell Rep*, 15, 481-9.
- WANG, Z., TROILO, P. J., WANG, X., GRIFFITHS, T. G., PACCHIONE, S. J., BARNUM, A. B., HARPER, L. B., PAULEY, C. J., NIU, Z., DENISOVA, L., FOLLMER, T. T., RIZZUTO, G., CILIBERTO, G., FATTORI, E., MONICA, N. L., MANAM, S. & LEDWITH, B. J. 2004b. Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther*, 11, 711-21.

- WARD, P., DEAN, F. B., O'DONNELL, M. E. & BERNIS, K. I. 1998. Role of the adenovirus DNA-binding protein in in vitro adeno-associated virus DNA replication. *J Virol*, 72, 420-7.
- WATT, G., LONG, G. W., PADRE, L. P., ALBAN, P., SANGALANG, R. & RANO, C. P. 1988. Chloroquine and quinine: a randomized, double-blind comparison of efficacy and side effects in the treatment of Plasmodium falciparum malaria in the Philippines. *Trans R Soc Trop Med Hyg*, 82, 205-8.
- WEI, F., ZHONG, S., MA, Z., KONG, H., MEDVEC, A., AHMED, R., FREEMAN, G. J., KROGSGAARD, M. & RILEY, J. L. 2013. Strength of PD-1 signaling differentially affects T-cell effector functions. *Proc Natl Acad Sci U S A*, 110, E2480-9.
- WEI, W. J., MU, S. R., HEINER, M., FU, X., CAO, L. J., GONG, X. F., BINDEREIF, A. & HUI, J. 2012. YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic Acids Res*, 40, 8622-36.
- WEI, Y., CHESNE, M. T., TERNS, R. M. & TERNS, M. P. 2015a. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res*, 43, 1749-58.
- WEI, Y., TERNS, R. M. & TERNS, M. P. 2015b. Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev*, 29, 356-61.
- WEINBERG, M. S., NICOLSON, S., BHATT, A. P., MCLENDON, M., LI, C. & SAMULSKI, R. J. 2014. Recombinant adeno-associated virus utilizes cell-specific infectious entry mechanisms. *J Virol*, 88, 12472-84.
- WEITZMAN, M. D., KYÖSTIÖ, S. R., KOTIN, R. M. & OWENS, R. A. 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc Natl Acad Sci U S A*, 91, 5808-12.
- WEN, L., ZHENG, Z. H., LIU, A. A., LV, C., ZHANG, L. J., AO, J., ZHANG, Z. L., WANG, H. Z., LIN, Y. & PANG, D. W. 2017. Tracking single baculovirus retrograde transportation in host cell via quantum dot-labeling of virus internal component. *J Nanobiotechnology*, 15, 37.
- WERTHER, W. A., GONZALEZ, T. N., O'CONNOR, S. J., MCCABE, S., CHAN, B., HOTALING, T., CHAMPE, M., FOX, J. A., JARDIEU, P. M., BERMAN, P. W. & PRESTA, L. G. 1996. Humanization of an anti-lymphocyte function-associated antigen (LFA)-1 monoclonal antibody and reengineering of the humanized antibody for binding to rhesus LFA-1. *J Immunol*, 157, 4986-95.
- WICKHAM, T. J. & NEMEROW, G. R. 1993. Optimization of growth methods and recombinant protein production in BTI-Tn-5B1-4 insect cells using the baculovirus expression system. *Biotechnol Prog*, 9, 25-30.
- WIEDENHEFT, B., ZHOU, K., JINEK, M., COYLE, S. M., MA, W. & DOUDNA, J. A. 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, 17, 904-12.
- WILDE, M., KLAUSBERGER, M., PALMBERGER, D., ERNST, W. & GRABHERR, R. 2014. Tnao38, high five and Sf9--evaluation of host-virus interactions in three different insect cell lines: baculovirus production and recombinant protein expression. *Biotechnol Lett*, 36, 743-9.
- WILSON, G. G., YOUNG, K. Y., EDLIN, G. J. & KONIGSBERG, W. 1979. High-frequency generalised transduction by bacteriophage T4. *Nature*, 280, 80-2.
- WINOCOUR, E. 1968. Further studies on the incorporation of cell DNA into polyoma-related particles. *Virology*, 34, 571-82.
- WISTUBA, A., KERN, A., WEGER, S., GRIMM, D. & KLEINSCHMIDT, J. A. 1997. Subcellular compartmentalization of adeno-associated virus type 2 assembly. *J Virol*, 71, 1341-52.
- WITTEWER, C. T., REED, G. H., GUNDRY, C. N., VANDERSTEEN, J. G. & PRYOR, R. J. 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem*, 49, 853-60.

- WOUTERS, J. A., ROMBOUTS, F. M., DE VOS, W. M., KUIPERS, O. P. & ABEE, T. 1999. Cold shock proteins and low-temperature response of *Streptococcus thermophilus* CNRZ302. *Appl Environ Microbiol*, 65, 4436-42.
- WU, K., MATTIOLI, M., MORSE 3RD, H. C. & DALLA-FAVERA, R. 2002. c-MYC activates protein kinase A (PKA) by direct transcriptional activation of the PKA catalytic subunit beta (PKA-C β) gene. *Oncogene*, 21, 7872-82.
- WU, P., XIAO, W., CONLON, T., HUGHES, J., AGBANDJE-MCKENNA, M., FERKOL, T., FLOTTE, T. & MUZYCZKA, N. 2000. Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *J Virol*, 74, 8635-47.
- WU, X., ZHANG, H., XING, Q., CUI, J., LI, J., LI, Y., TAN, Y. & WANG, S. 2014. PD-1(+) CD8(+) T cells are exhausted in tumours and functional in draining lymph nodes of colorectal cancer patients. *Br J Cancer*, 111, 1391-9.
- WU, Y., LIANG, D., WANG, Y., BAI, M., TANG, W., BAO, S., YAN, Z., LI, D. & LI, J. 2013. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell*, 13, 659-62.
- WU, Y., ZHOU, H., FAN, X., ZHANG, Y., ZHANG, M., WANG, Y., XIE, Z., BAI, M., YIN, Q., LIANG, D., TANG, W., LIAO, J., ZHOU, C., LIU, W., ZHU, P., GUO, H., PAN, H., WU, C., SHI, H., WU, L., TANG, F. & LI, J. 2015. Correction of a genetic disease by CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells. *Cell Res*, 25, 67-79.
- WU, Z., YANG, H. & COLOSI, P. 2010. Effect of genome size on AAV vector packaging. *Mol Ther*, 18, 80-6.
- WYVEKENS, N., TOPKAR, V. V., KHAYTER, C., JOUNG, J. K. & TSAI, S. Q. 2015. Dimeric CRISPR RNA-guided FokI-dCas9 nucleases directed by truncated gRNAs for highly specific genome editing. *Hum Gene Ther*, 26, 425-31.
- XIAO, P. J., MITCHELL, A. M., HUANG, L., LI, C. & SAMULSKI, R. J. 2016. Disruption of microtubules post-virus entry enhances adeno-associated virus vector transduction. *Hum Gene Ther*, 27, 309-24.
- XIAO, P. J. & SAMULSKI, R. J. 2012. Cytoplasmic trafficking, endosomal escape, and perinuclear accumulation of adeno-associated virus type 2 particles are facilitated by microtubule network. *J Virol*, 86, 10462-73.
- XIAO, X., LI, J. & SAMULSKI, R. J. 1996. Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. *J Virol*, 70, 8098-108.
- XIAO, X., LI, J. & SAMULSKI, R. J. 1998. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J Virol*, 72, 2224-32.
- XIE, Q., BU, W., BHATIA, S., HARE, J., SOMASUNDARAM, T., AZZI, A. & CHAPMAN, M. S. 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc Natl Acad Sci U S A*, 99, 10405-10.
- XU, H., XIAO, T., CHEN, C. H., LI, W., MEYER, C. A., WU, Q., WU, D., CONG, L., ZHANG, F., LIU, J. S., BROWN, M. & LIU, X. S. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Res*, 25, 1147-57.
- XU, J., LIAN, W., JIA, Y., LI, L. & HUANG, Z. 2017. Optimized guide RNA structure for genome editing via Cas9. *Oncotarget*, 8, 94166-71.
- XU, X., CHEN, Y., ZHAO, Y., LIU, X., DONG, B., JONES, I. M. & CHEN, H. 2013. Baculovirus superinfection: a probable restriction factor on the surface display of proteins for library screening. *PLoS One*, 8, e54631.
- XUE, C., ZHU, Y., ZHANG, X., SHIN, Y. K. & SASHITAL, D. G. 2017. Real-time observation of target search by the CRISPR surveillance complex cascade. *Cell Rep*, 21, 3717-27.
- XUE, J., QIAO, N., ZHANG, W., CHENG, R. L., ZHANG, X. Q., BAO, Y. Y., XU, Y. P., GU, L. Z., HAN, J. D. & ZHANG, C. X. 2012. Dynamic interactions between *Bombyx mori*

- nucleopolyhedrovirus and its host cells revealed by transcriptome analysis. *J Virol*, 86, 7345-59.
- YALKINOGLU, A. O., HEILBRONN, R., BÜRKLE, A., SCHLEHOFER, J. R. & ZUR HAUSEN, H. 1988. DNA amplification of adeno-associated virus as a response to cellular genotoxic stress. *Cancer Res*, 48, 3123-9.
- YANG, L., GUELL, M., NIU, D., GEORGE, H., LESHA, E., GRISHIN, D., AACH, J., SHROCK, E., XU, W., POZI, J., CORTAZIO, R., WILKINSON, R. A., FISHMAN, J. A. & CHURCH, G. 2015. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science*, 350, 1101-4.
- YANG, Q., CHEN, F. & TREMPER, J. P. 1994. Characterization of cell lines that inducibly express the adeno-associated virus Rep proteins. *J Virol*, 68, 4847-56.
- YANG, W. H. & BLOCH, D. B. 2007. Probing the mRNA processing body using protein macroarrays and "autoantigenomics". *RNA*, 13, 704-12.
- YAO, T., SONG, L., JIN, J., CAI, Y., TAKAHASHI, H., SWANSON, S. K., WASHBURN, M. P., FLORENS, L., CONAWAY, R. C., COHEN, R. E. & CONAWAY, J. W. 2008. Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the proteasome and in the Ino80 chromatin-remodeling complex. *Mol Cell*, 31, 909-17.
- YE, L., WANG, J., BEYER, A. I., TEQUE, F., CRADICK, T. J., QI, Z., CHANG, J. C., BAO, G., MUENCH, M. O., YU, J., LEVY, J. A. & KAN, Y. W. 2014. Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. *Proc Natl Acad Sci U S A*, 111, 9591-6.
- YOSEF, I., GOREN, M. G. & QIMRON, U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res*, 40, 5569-76.
- YU, Q., XIONG, Y., GAO, H., LIU, J., CHEN, Z., WANG, Q. & WEN, D. 2015. Comparative proteomics analysis of *Spodoptera frugiperda* cells during *Autographa californica* multiple nucleopolyhedrovirus infection. *Virology*, 12, 115.
- YU, X., LIANG, X., XIE, H., KUMAR, S., RAVINDER, N., POTTER, J., DE MOLLERAT DU JEU, X. & CHESNUT, J. D. 2016. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnol Lett*, 38, 919-29.
- YUEN, G., KHAN, F. J., GAO, S., STOMMEL, J. M., BATCHELOR, E., WU, X. & LUO, J. 2017. CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. *Nucleic Acids Res*, 45, 12039-53.
- YUEN, K. S., CHAN, C. P., WONG, N. H., HO, C. H., HO, T. H., LEI, T., DENG, W., TSAO, S. W., CHEN, H., KOK, K. H. & JIN, D. Y. 2015. CRISPR/Cas9-mediated genome editing of Epstein-Barr virus in human cells. *J Gen Virol*, 96, 626-36.
- ZABOIKIN, M., ZABOIKINA, T., FRETER, C. & SRINIVASAKUMAR, N. 2017. Non-Homologous End Joining and Homology Directed DNA Repair Frequency of Double-Stranded Breaks Introduced by Genome Editing Reagents. *PLoS One*, 12, e0169931.
- ZAIDI, A. U., MCDONOUGH, J. S., KLOCKE, B. J., LATHAM, C. B., KORSMEYER, S. J., FLAVELL, R. A., SCHMIDT, R. E. & ROTH, K. A. 2001. Chloroquine-induced neuronal cell death is p53 and Bcl-2 family-dependent but caspase-independent. *J Neuropathol Exp Neurol*, 60, 937-45.
- ZAMORE, P. D., TUSCHL, T., SHARP, P. A. & BARTEL, D. P. 2000. RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101, 25-33.
- ZASEDATELEVA, O. A., KRYLOV, A. S., PROKOPENKO, D. V., SKABKIN, M. A., OVCHINNIKOV, L. P., KOLCHINSKY, A. & MIRZABEKOV, A. D. 2002. Specificity of mammalian Y-box binding protein p50 in interaction with ss and ds DNA analyzed with generic oligonucleotide microchip. *Journal of Molecular Biology*, 324, 73-87.

- ZELTNER, N., KOHLBRENNER, E., CLEMENT, N., WEBER, T. & LINDEN, R. M. 2010. Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. *Gene Ther*, 17, 872-9.
- ZENG, Y., YI, R. & CULLEN, B. R. 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proceedings of the National Academy of Sciences*, 100, 9779-84.
- ZETSCHKE, B., GOOTENBERG, J. S., ABUDAYYEH, O. O., SLAYMAKER, I. M., MAKAROVA, K. S., ESSLETZBICHLER, P., VOLZ, S. E., JOUNG, J., VAN DER OOST, J., REGEV, A., KOONIN, E. V. & ZHANG, F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163, 759-71.
- ZHANG, Y., HEIDRICH, N., AMPATTU, B. J., GUNDERSON, C. W., SEIFERT, H. S., SCHOEN, C., VOGEL, J. & SONTHEIMER, E. J. 2013. Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol Cell*, 50, 488-503.
- ZHANG, Y. & HEMPELMANN, E. 1987. Lysis of malarial parasites and erythrocytes by ferriprotoporphyrin IX-chloroquine and the inhibition of this effect by proteins. *Biochem Pharmacol*, 36, 1267-73.
- ZHANG, Y. & SKOLNICK, J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57, 702-10.
- ZHANG, Y. & SKOLNICK, J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, 33, 2302-9.
- ZHANG, Y. F., HOMER, C., EDWARDS, S. J., HANANEIA, L., LASHAM, A., ROYDS, J., SHEARD, P. & BRAITHWAITE, A. W. 2003. Nuclear localization of Y-box factor YB1 requires wild-type p53. *Oncogene*, 22, 2782-94.
- ZHANG, Z., SCHWARTZ, S., WAGNER, L. & MILLER, W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol*, 7, 203-14.
- ZHAO, D., FENG, X., ZHU, X., WU, T., ZHANG, X. & BI, C. 2017. CRISPR/Cas9-assisted gRNA-free one-step genome editing with no sequence limitations and improved targeting efficiency. *Sci Rep*, 7, 16624.
- ZHAO, Y., ANDO, K., OKI, E., IKAWA-YOSHIDA, A., IDA, S., KIMURA, Y., SAEKI, H., KITAO, H., MORITA, M. & MAEHARA, Y. 2014. Aberrations of BUBR1 and TP53 gene mutually associated with chromosomal instability in human colorectal cancer. *Anticancer Res*, 34, 5421-7.
- ZHEN, H. N., ZHANG, X., BU, X. Y., ZHANG, Z. W., HUANG, W. J., ZHANG, P., LIANG, J. W. & WANG, X. L. 1999. Expression of the simian virus 40 large tumor antigen (Tag) and formation of Tag-p53 and Tag-pRb complexes in human brain tumors. *Cancer*, 86, 2124-32.
- ZHONG, Z., WU, Z., HAN, L. & CHEN, J. 2017. Novel mutations in CRYGC are associated with congenital cataracts in Chinese families. *Sci Rep*, 7, 189.
- ZHU, P., WU, F., MOYSESON, J., ZHANG, H., HE, T. C. & WU, W. S. 2017. CRISPR/Cas9-Mediated Genome Editing Corrects Dystrophin Mutation in Skeletal Muscle Stem Cells in a Mouse Model of Muscle Dystrophy. *Mol Ther Nucleic Acids*, 7, 31-41.
- ZHU, W., LEI, R., LE DUFF, Y., LI, J., GUO, F., WAINBERG, M. A. & LIANG, C. 2015. The CRISPR/Cas9 system inactivates latent HIV-1 proviral DNA. *Retrovirology*, 12, 22.
- ZHU, Y. & QI, Y. 1999. Construction of transposon-mediated baculovirus vector and expression of green fluorescent protein in insect cells and larvae. *Chinese Science Bulletin*, 44, 158-63.
- ZINDER, N. D. & LEDERBERG, J. 1952. Genetic exchange in *Salmonella*. *J Bacteriol*, 64, 679-99.
- ZOLOTUKHIN, S., BYRNE, B. J., MASON, E., ZOLOTUKHIN, I., POTTER, M., CHESNUT, K., SUMMERFORD, C., SAMULSKI, R. J. & MUZYCZKA, N. 1999. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther*, 6, 973-85.

- ZOLOTUKHIN, S., POTTER, M., ZOLOTUKHIN, I., SAKAI, Y., LOILER, S., FRAITES, T. J. J., CHIDO, V. A., PHILLIPSBERG, T., MUZYCZKA, N., HAUSWIRTH, W. W., FLOTTE, T. R., BYRNE, B. J. & SNYDER, R. O. 2002. Production and purification of serotype 1, 2, and 5 recombinant adeno-associated viral vectors. *Methods Enzymol*, 28, 158-67.
- ZUCKERKANDL, E. & PAULING, L. B. 1965. *Evolutionary divergence and convergence in proteins*, New York, NY, Academic Press.