

Northumbria Research Link

Citation: Nguyen, Dinh Dung (2022) Targeted modification of major histocompatibility complex class 1 molecules via CRISPR-CAS9. Doctoral thesis, Northumbria University.

This version was downloaded from Northumbria Research Link:
<https://nrl.northumbria.ac.uk/id/eprint/51269/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>



**Northumbria
University**
NEWCASTLE



UniversityLibrary

**TARGETED MODIFICATION OF MAJOR
HISTOCOMPATIBILITY COMPLEX CLASS I
MOLECULES VIA CRISPR-CAS9**

D D NGUYEN

PhD

2022



**Northumbria
University**
NEWCASTLE

**TARGETED MODIFICATION OF MAJOR
HISTOCOMPATIBILITY COMPLEX CLASS I
MOLECULES VIA CRISPR-CAS9**

DINH-DUNG NGUYEN

A thesis submitted in partial fulfilment of the requirements of the
University of Northumbria at Newcastle for the degree of Doctor
of Philosophy

Research undertaken in the Department of Applied Sciences,
Faculty of Health & Life Sciences

Abstract

Major Histocompatibility Complex (MHC) class I molecules present peptides derived from intracellular proteins for cytotoxic T cell recognition against virally infected cells/tumor cells and inducing adaptive immunity. MHC-I complexes are comprised of a heavy chain, the light chain beta-2 microglobulin (B2M) and an optimized peptide. Recently, Complex of Clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein Cas9 has rapidly become a powerful tool for gene-editing experiments due to its high efficiency, simplicity, and versatility. This thesis proposes two approaches to apply CRISPR/Cas9-mediated knock-in methodology to modify MHC-I complexes: (i) to stably introduce specific peptides on MHC-I molecules through CRISPR-mediated insertion within the *B2M* locus; (ii) target single amino acid residues of the Ankylosing Spondylitis (AS)-associated HLA-B27 molecule proposed to contribute to its unusual biochemical folding characteristics.

CRISPR/Cas9 was successfully employed to knock in HLA-A2 specific viral peptides, derived from SARS-CoV-2 virus, containing linker sequences at the genomic *B2M* locus. Immunoblotting revealed that the gene edited *B2M* locus facilitates the production of peptide linked β 2m (PLB) protein with the expected size. Furthermore, flow cytometry analysis demonstrated gene-edited cells presented the PLB molecules together with endogenous peptide-specific HLA-A2 heavy chain at the cell surface.

CRISPR/Cas9 was also harnessed to introduce point mutations into the disease-associated HLA-B*27:05 allele. The absence of Cys⁶⁷ can potentially reduce the misfolding events of HLA-B27 molecules and still maintain the presence of this heavy chain at the cell surface. Additionally, HLA-B*27:05 have been transformed into HLA-B*27:09, a non-AS-associated subtype, by replacing aspartic acid (Asp)¹¹⁶ with a histidine (His) using CRISPR/Cas9.

This is the first study exploiting gene-editing technology for modification of MHC-I complexes. While stable expression of PLB is a promising approach for development of cellular vaccines, alterations of HLA class I sequences can be considered as a potential therapeutic application for some HLA-associated diseases.

Contents

Abstract.....	i
Contents.....	ii
List of Figures	vii
List of Tables	xi
Abbreviation	xii
Acknowledgement.....	xvii
Author's declaration.....	xviii
Chapter 1 Introduction.....	20
1.1 Major Histocompatibility Complex	20
1.2 MHC class I.....	25
1.2.1 Heavy chain of MHC-I.....	25
1.2.2 Polymorphisms of MHC-I.....	27
1.2.3 The Light Chain of MHC-I; Beta-2 Microglobulin	29
1.2.4 MHC-I Protein Folding and Assembly.....	32
1.3 MHC-I specific antigens	35
1.3.1 Selection and presentation of MHC-I specific antigens	35
1.3.2 Manipulation of MHC-I peptide for clinical application	38
1.4 Ankylosing Spondylitis and <i>HLA-B27</i> gene	47
1.4.1 Clinical features of Ankylosing Spondylitis.....	47
1.4.2 Pathogenesis of AS	49

1.4.3	Key structural features of HLA-B27 molecules	59
1.5	Gene-editing Technology.....	63
1.5.1	Mechanism of gene editing.....	64
1.5.2	Classification of Gene-editing Technology.....	68
1.5.2	CRISPR-based technologies	72
1.6	Application of Gene-editing Technology in Immunology	86
1.6.1	General application in immunology.....	86
1.6.2	Targeting the MHC-I Complex by CRISPR/Cas9 technology	90
1.7	Thesis rationale and objectives	94
Chapter 2 Materials and Methods		97
2.1	Cell culture	97
2.1.1	Culture Cell.....	97
2.1.2	Mammalian cell transfection	98
2.1.3	Flow cytometry analysis.....	101
2.1.4	Single cell cloning.....	102
2.2	Molecular biology	104
2.2.1	Bacterial Strains	104
2.2.2	Bacterial transformations	104
2.2.3	Extraction of genomic DNA, RNA and plasmid extraction	104
2.2.4	cDNA synthesis	107
2.2.5	Polymerase Chain Reaction (PCR)	108

2.2.6	Agarose gel electrophoresis	109
2.2.7	PCR purification.....	109
2.2.8	Cell lysis	111
2.2.9	Immunoblotting	111
2.2.10	Restriction enzyme digestion and ligase	113
2.2.11	Plasmid transformation.....	114
2.2.10	Site-directed mutagenesis.....	114
2.2.11	Generating stable mammalian expression cell lines using the Flp-in system	115
2.2.12	<i>In vitro</i> cleavage assay.....	116
2.2.13	<i>In vivo</i> cleavage assay	117
2.2.14	Next-Generation Sequencing	118
2.3	Bioinformatics and statistical analysis.....	119
2.2.1	Plasmid map and sequence retrieval.....	119
2.3.2	Data analysis and graphic illustrations	119
Chapter 3 Rescue of MHC-I Expression By Peptide-linked B2M.....		121
3.1	Chapter aims & objectives	121
3.2	Results	122
3.3.1	Strategy of Peptide-linker-B2M plasmid construction	122
3.2.2	Selection of <i>HLA-A*02:01</i> specified peptides	126
3.3.3	Generation of pCR3.1-peptide-linker-B2M plasmids	130

3.3.4	Generating a test B2M Negative HEK cell line	136
3.3.5	Expression of PLB in B2M-deficient HEK cells	144
3.3	Discussion.....	149
Chapter 4 Generating peptide-linked B2M using CRISPR/Cas9-mediated knock-in.....		154
4.1	Chapter aims & objectives	154
4.2	Results 157	
4.2.1	CRISPR/Cas9 system targeting exon 1 of B2M	157
4.2.2	Optimization of HDR to insert peptide and linkers into the endogenous <i>B2M</i> locus	161
4.2.3	Expression of PLB-E629–38-(GGGGS) ₃ proteins in HEK cell.....	165
4.2.4	Expression of PLB-N325–333-(GGGGS) ₃ on HEK.....	170
4.2.5	Expression of PLB-S269–277-G ₆ (SGG) ₃ by HEK cells.....	180
4.2.6	Evaluate off-target effects on gene-edited cell lines	189
4.3	Discussion.....	192
Chapter 5 Modification of the <i>HLA-B27</i> gene by CRISPR technology.....		198
5.1	Chapter aims & objectives	198
5.2	Results 200	
5.2.1	Generation of HLA-B27 mutants C67S, D116H and C101S by Site Directed Mutagenesis	200
5.2.2	Investigating C67S, D116H and C101S.B27 expressing cell lines	206
5.2.3	CRISPR/Cas9 system targeting HLA-B*27:05.....	213

5.2.4	Selection of cell lines for HDR	218
5.2.6	HDR of C67S, C101S and D116H using gRNA and ssODN	222
5.2.7	Cloning of gene-edited cells	226
5.2.8	Investigating gene edited HLA-B27 expression	229
5.2.9	Evaluating off-target effects on gene-edited cell lines	233
5.3	Discussion.....	236
Chapter 6	Conclusions	241
6.1	General discussion.....	241
6.2	Limitations and future work	245
6.3	Final remarks	250
Appendix 1:	List of chemicals.....	255
Appendix 2:	List of equipment.....	259
Appendix 3:	Plasmid map and sequence retrieval	260
Appendix 4:	Data analysis.....	265
Appendix 5:	Primers for Polymerase Chain Reaction.....	277
REFERENCES	280

List of Figures

Figure 1.1 Genomic landscape of the MHC.....	22
Figure 1.2. Structure of major histocompatibility complex (MHC) class I.	26
Figure 1.3. Pairwise comparison of <i>HLA-A</i> , <i>-B</i> and <i>-C</i> alleles.....	28
Figure 1.4. Different types of vaccines and their various immunogenicity mechanisms.	41
Figure 1.5. Spinal joint of healthy and Ankylosing Spondylitis individuals.	47
Figure 1.6. Proposed HLA-B27-related-pathogenic mechanisms of Ankylosing Spondylitis.	53
Figure 1.8: Structural features of HLA-B*27:05 and HLA-B*27:09.	61
Figure 1.9. DNA double-strand break repair via non-homologous end joining and homologous recombination pathways.	66
Figure 1.10. Schematic diagram of four gene-editing endonuclease systems.	68
Figure 1.11. Schematic highlighting components of sgRNA and Cas9 nuclease.	73
Figure 1.12. Property Comparison of Natural CRISPR/Cas9 Systems.	75
Figure 1.13. Schematic diagram of features of CRISPR/Cas9 platforms including plasmids, mRNAs and RNPs.	79
Figure 1.14. Cas9 RNP delivery routes and their features.....	82
Figure 1.15. The application of CRISPR/Cas9 in immunotherapy.....	88
Figure 1.16. Strategies of CRISPR/Cas9-mediated disruption of HLA genes for generation of universal iPSC with enhanced immune compatibility.	93
Figure 2.1. Workflow of AMPure XP Bead Purification.....	110
Figure 3.1. Construction of single chain trimeric molecule and peptide linked B2M molecules.	124
Figure 3.2. Cloning of B2M cDNA into pCR3.1 plasmid.	126
Figure 3.3. Generation of <i>pCR3.1-E629-38-(GGGGS)3-B2M</i> and <i>pCR3.1-N325-333-(GGGGS)3-B2M</i>	132
Figure 3.4. Generation of <i>pCR3.1-S269-277-(GGGGS)3-B2M</i> and <i>pCR3.1-S269-277-G6(SGG)3-B2M</i> containing plasmids.	135
Figure 3.5. Selection of sgRNAs targeting <i>B2M</i>	137
Figure 3.6. <i>In vitro</i> cleavage assay of Cas9 RNPs targeting <i>B2M</i>	138

Figure 3.7. <i>In vivo</i> cleavage result of Cas9 RNPs targeting B2M.....	140
Figure 3.8. Investigation of surface MHC-I expression by gene-edited K562 cells.....	141
Figure 3.9. Identification of a homologous B2M knock-out HEK cell line.....	143
Figure 3.10. Strategy to evaluate function of pPLB plasmids.....	144
Figure 3.11. Flow-cytometry analysis of B2M-deficient cell transfected with pPLB plasmids.	146
Figure 3.12. Immunoblotting of B2M and PLB proteins.....	148
Figure 4.1. Antigen presentation pathways of MHC-I, SCT and MHC-I-PLB complex.....	156
Figure 4.2. Design of sgRNA targeting the end of B2M.exon_1.	157
Figure 4.3. Functional assays of sgRNA.B2M.1.3.	159
Figure 4.4. INDEL spectrum of sgRNA.B2M.1.3 targeting the <i>B2M</i> locus.	160
Figure 4.5. Design of donor templates for HDR of PLB.	161
Figure 4.6. Detection of HDR of PLB insertion in K562 and HEK cells.....	163
Figure 4.7. Evaluating insertion of peptide and linker into the endogenous B2M allele on HEK cells.	164
Figure 4.8. Single-cell isolation of monoclonal HEK cells carrying E6 ₂₉₋₃₈ /(GGGGG) ₃	166
Figure 4.9. Flow-cytometry analysis of MHC-I expression by HEK-E6 ₂₉₋₃₈ /(GGGGG) ₃ cells.....	168
Figure 4.10. Detection of B2M and PLB by HEK-E6 ₂₉₋₃₈ -(GGGGG) ₃ cells.....	169
Figure 4.11. HDR results of PLB- N ₃₂₅₋₃₃₃ -(GGGGG) ₃ insertion in HEK cells	171
Figure 4.12. Single-cell isolation for monoclonal HEK-N ₃₂₅₋₃₃₃ /(GGGGG) ₃ clones from polyclonal gene-edited populations.	172
Figure 4.13. Characterization of surface MHC-I expression on HEK-N ₃₂₅₋₃₃₃ -(GGGGG) ₃ cells.	174
Figure 4.14. Immunoblotting of B2M protein on HEK-N ₃₂₅₋₃₃₃ /(GGGGG) ₃ cells.	175
Figure 4.15. Illustration of intron motif on N ₃₂₅₋₃₃₃ -(GGGGG) ₃	177
Figure 4.16. Detection of PLB mRNA expressed by gene-edited cells.	178
Figure 4.17. Removal of potential intronic splice/acceptor motifs within the G ₆ (SGG) ₃ linker.	180
Figure 4.18. Results from the HDR of PLB- S ₂₆₉₋₂₇₇ -G ₆ (SGG) ₃ sequence insertion into HEK cells.....	181
Figure 4.19. Single-cell HEK-S ₂₆₉₋₂₇₇ -G ₆ (SGG) ₃ clones assessed for HDR insertion of S269-277-G ₆ (SGG) ₃ sequence.	183

Figure 4.20. Flow-cytometry analysis of MHC-I expression by HEK-S ₂₆₉₋₂₇₇ -G ₆ (SGG) ₃ cells.	185
Figure 4.21. Immunoblotting of B2M in HEK- S ₂₆₉₋₂₇₇ -G ₆ (SGG) ₃ clonal cell populations.	187
Figure 4.22. On- and Off-target analysis of HEK cells treated with Cas9 RNP with sgRNA.B2M.1.3.	191
Figure 5.1. Nucleotide substitutions introducing the desired amino acid changes and restriction enzyme sites within the HLA-B*27:05 sequence.	201
Figure 5.2. Specific primers for amplifying cDNA of B27 from the pcDNA™5/B27/FRT/V5 plasmid.....	202
Figure 5.3. Site-directed mutagenesis of pcDNA™5/B27/FRT/V5 introducing C67S, D116H and C101S changes.	204
Figure 5.4. Confirmation of mutated-HLA-B27 integration into HeLa-H2Z.....	206
Figure 5.5. Flow-cytometry analysis of HeLa cell lines expressing HLA-B27 and B27 mutants.	208
Figure 5.6. Immunoblotting analysis of MHC-I expression on HeLa cell lines.	211
Figure 5.7. Selection of sgRNAs targeting the HLA-B*27:05 allele.	214
Figure 5.8. <i>In vitro</i> cleavage assay of Cas9 RNP targeting HLA-B27 cDNA	216
Figure 5.9. <i>In vivo</i> T7EI cleavage result of Cas9 RNPs targeting HLA-B27.	218
Figure 5.10. Design of sgRNA and donor template for HDR targeting exon 1 of the B2M locus.	219
Figure 5.11. Evaluation of HDR in HeLa, HEK and K562 cell lines.	221
Figure 5.12. Design of donor templates built for homology-directed repair to introduce single amino acid changes within the HLA-B27 sequence.	223
Figure 5.13. Genome editing by transfection of B27.D116H.ssODN and Cas9/sgRNA RNP targeting the HLA- B27 cDNA expressed by HeLa-B27 cell lines.	225
Figure 5.14. Genome editing by transfection of ssODNs and Cas9/sgRNA RNPs targeting HLA-B27 in K562- HLA-B27 cell line.	226
Figure 5.15. Single-cell isolation of monoclonal K562-HLA-B27 cells containing target mutations.	227
Figure 5.16. Confirmation of nucleotide substitutions in monoclonal gene-edited K562-HLA-B27 cells.	228
Figure 5.17. Characterisation of surface MHC-I expression by gene-edited K562-HLA-B27 cells.....	230
Figure 5.18. Immunoblotting of HLA-I heavy chain expressed by gene-edited cells.	232

Figure 5.19. On-target and off-target analysis of the RNP of Cas9 protein and sg272F on K562-HLA-B27 cells
..... 235

Figure 6.1. CRISPR-mediated modifications of MHC-I complexes and their potential applications..... 252

List of Tables

Table 1.1 Examples of different functions and applications of B2M.....	31
Table 1.2. Summary of studies using MHC-I single-chain trimers to induce CTL responses.....	44
Table 1.3. Amino acid differences between 6 of the most frequent HLA-B27 allotypes and non-AS associated HLA-B*27:06 and 09 subtypes.	60
Table 1.4 Characteristics of viral vectors in clinical application.....	77
Table 1.5. Summary of studies using CRISPR/Cas9 technology to disrupt the expression of MHC-I molecules.	90
Table 2.1. Information of cell lines and growth medium.....	98
Table 2.2. Amount of jetPRIME reagents and DNA according to the cell culture vessel per well.....	99
Table 2.3. Amount of CRISPRMAX reagents, Cas9 nuclease and sgRNA according to the cell culture vessel per well.....	100
Table 2.4. Nucleofection conditions and amounts of HDR enhancers, Cas9 nuclease, sgRNA and ssODN according to cell lines employed.....	101
Table 2.5. List of antibodies used for flow-cytometry.....	102
Table 2.6. Components of PCR.....	108
Table 2.7. The list of antibodies used for immunoblotting.....	113
Table 2.8. Components of restriction digest.....	114
Table 3.1. Summary of HLA-A*02:01 specified peptides and linkers used to develop SCT molecules and stimulate CD8+ cytotoxic T cells.....	127
Table 3.2. Predicted specified HLA-A*02:01 binding peptides from S and N proteins of SARS-CoV-2.....	128
Table 3.3. Summary of selected HLA-A*02:01 specified peptides and linker for designing construct of Peptide-linked B2M (PLB).....	130
Table 4.1. On-target and Off-target sites for RNP of Cas9 protein and sgRNA.B2M.1.3.....	189
Table 5.1. On-target and off-target sites for the RNP of Cas9 protein and sg272F.....	233

Abbreviation

Abbreviation	Description
AAV	Adeno-associated Virus
ALL	Acute Lymphoblastic Leukemia
APC	Antigen Presenting Cell
AS	Ankylosing Spondylitis
ATF4	Activating transcription factor 4
ATF6	Activating transcription factor 6
ATF6 α	activating transcription factor 6 alpha
ATM	ataxia telangiectasia mutated protein
ATP	Adenosine 5'-triphosphate
AV	Adenovirus
B2M	Beta-2-Microglobulin
BCA	Bicinchoninic acid
BCR	B cell receptor
CAR-T	Chimeric antigen receptor T cell
CD1	Cluster of Differentiation 1
CDC	Conventional Dendritic cell
CMA	Chromosomal microarray analysis
CNX	Calnexin
COVID-19	Coronavirus disease 2019
CPI	Complete protease inhibitors
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
CrRNA	CRISPR RNA
CRT	Calreticulin
CTL	Cytotoxic T lymphocyte
DC	Dendritic Cell

DGK	Diacylglycerol kinase
dLNPs	Dendrimer-based lipid nanoparticles
DMD	Duchenne Muscular Dystrophy
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	DNA Double-strand Breaks
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal Growth Factor Receptor
ELISPOT	Enzyme-linked immunosorbent spot
ER	Endoplasmic Reticulum
ERAP1	Endoplasmic reticulum aminopeptidase 1
ERAP2	Endoplasmic reticulum aminopeptidase 2
EVIR	Vesicle-internalizing receptor
FACS	Fluorescence-activated Cell Sorting
FBS	Fetal Bovine Serum
FCRn	Neonatal Fc Receptor
FRT	Flippase recognition target
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GWAS	Genome-wide association studie
HC	Heavy Chain
HCMV	Human Cytomegalovirus
HDR	Homology-Directed Repair
HFE	High FE ²⁺
HLA	Human Leukocyte Antigen
HPRT	Human Hypoxanthine-guanine phosphoribosyl transferase

HPV	Human Papillomaviruses
HR	Homologous recombination
HSPC	Hematopoietic stem and Progenitor cell
HSV	Herpes simplex virus
ICI	Immune checkpoint inhibitors
IDLV	Integration-defective Lentivirus
IFN- γ	Interferon gamma
iPSC	Induced pluripotent stem cell
IRE1	Inositol-requiring protein 1
IRF	Interferon regulatory factor
IVT	In vitro
IVV	In vivo
KIR	Killer Cell Immunoglobulin-Like Receptors
KIR3DL1	Killer cell immunoglobulin-like receptors 1
KIR3DL2	Killer cell immunoglobulin-like receptors 2
KO	Knock-out
LB	Luria Broth
LIG4	DNA ligase IV
LV	Lentivirus
MegNs	Meganucleases
MFI	Mean Fluorescence Intensities
MHC	Major Histocompatibility Complex
N	Nucleocapsid
NCBI	National Center for Biotechnology Information
NEM	N-methylmaleimide
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cell
NGS	Next-generation sequencing
NHEJ	Non-Homologous End Joining

NK	Natural Killer
NKIR	Natural Killer Inhibitory Receptor
NLS	Nuclear Localization Signal
PAGE	Polyacrylamide gel electrophoresis
PAK	P21-activated kinase
PAM	Protospacer Adjacent Motif
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PDI	Protein Disulfide Isomerases
PD-L1	Programmed death-ligand 1
PEG	Polyethylene Glycol
PEI	Polyethyleneimine
PERK	Protein kinase RNA-like ER kinase
PF33	PEI-derived fluorinated polymer
PKU	Phenylketonuria
PLC	Peptide Loading Complex
PMSF	Phenylmethanesulphonyl fluoride
PsA	Psoriatic Spondylarthritis
ReA	Reactive Arthritis
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RNPs	Ribonucleoproteins
RPA	Replication protein A
S	Spike
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SCT	Single Chain Trimer
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulphate

SEM	Standard Error of Mean
SgRNA	Single-guide RNA
SNP	Single nucleotide polymorphism
SpA	Spondyloarthropathy
ssDNA	Single-stranded DNA
TAA	Tumor-associated antigen
TABPR	TAP-binding protein related
TAE	Tris-acetate-EDTA
TALEN	Transcription Activator-like Effector Nucleases
TAP	Transporter Associated with Antigen Processing
TEMED	N,N,N',N'-Tetramethyl ethylenediamine
TGFBR2	Transforming growth factor beta receptor II
TGF- β	Transforming Growth Factor Beta
TLR	toll-like receptor
ToIDC	Tolerance Dendritic cell
TracrRNA	Trans-activating crRNA
UPR	Unfolded Protein Response
WGS	Whole Genome Sequencing
XBP1	X-box binding protein 1
XLF	XRCC4-like factor
XRCC4	X-ray cross-complementing protein 4
YTHDF1	YTH N6-Methyladenosine RNA Binding Protein 1

Acknowledgement

This project is supported by Vinmec Healthcare System.

I would like to thank my supervisor, Dr Antony Antoniou, for all his time and continued support over the last 4 years of my PhD. His encouragement and trust are the keys to helping me discover and promote my abilities in my research path. I would also like to thank Dr Sterghios Moschos and Dr Simon Powis for their guidance and supervision. Personal thanks go to Dr Nguyen Thanh Liem for always being an inspiration and a model for me to pursue my passion.

Furthermore, I owe particular thanks to all my colleagues and staff at both Northumbria University and Vinmec Healthcare System for their moral support throughout the course of my studies. Thanks to all of my friends, especially members of the Vietnamese Badminton Society in Newcastle, for being with me on my journey over the years.

Finally, I would like to send the deepest gratitude to my family for their endless love, support, inspiration and caring. I would like to dedicate this thesis to my Dad and Mom, Mr Nguyen Dinh Ty and Mrs Nguyen Thi Kim Dung, the greatest people I've ever known. Thank you for all the love and support.

Vietnamese translation: Cuối cùng, tôi xin gửi lời cảm ơn sâu sắc nhất tới gia đình vì tình yêu thương, sự ủng hộ, nguồn cảm hứng và sự quan tâm vô bờ bến của mọi người. Tôi muốn dành tặng luận văn này cho Ba Mẹ tôi, Ông Nguyễn Đình Tý và Bà Nguyễn Thị Kim Dung, những người vĩ đại nhất mà tôi từng biết. Cảm ơn ba mẹ vì tất cả tình yêu và sự ủng hộ dành cho con.

Author's declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas, and contributions from the work of others.

Any ethical clearance for the research presented in this commentary has been approved. Approval has been sought and granted through the Researcher's submission to Northumbria University's Ethics Online System [Submission Ref: 21627] on [09/07/2020].

I declare that the Word Count of this Thesis is 44849 words

Name: Dinh Dung Nguyen

Date: 20/07/2022

CHAPTER 1
INTRODUCTION

Chapter 1 Introduction

1.1 Major Histocompatibility Complex

The Major Histocompatibility Complex (MHC), is a gene locus with thousands of genes encoding for important proteins involved in both the innate and adaptive immune response, was first described by British immunologist Peter Gorer in 1936 [1]. In 1948, George Snell published a pioneering paper introducing the concept of histocompatibility (H) antigens, and the genetic factors which played a central role in graft rejection of tumours transplanted across different mouse strains [2]. Snell found that these genetic factors were located within a specific chromosomal area, which was later to be known as the MHC. In 1958, Jean Dausset identified the very first one of these genetic systems, Human Leukocyte Antigen (*HLA*)-A*2 which were initially named MAC the three-letters of the first name of his three volunteers [3]. In the 1960s, Baruj Benacerraf and his colleagues demonstrated that the MHC genes had several central functions in immune responses [4]. Thanks to their contributions, Baruj Benacerraf, Jean Dausset and George Snell were awarded jointly The Nobel Prize in Physiology or Medicine 1980 for "their discoveries concerning genetically determined structures on the cell surface that regulate immunological reactions" (Nobelprize.org). These scientists laid the foundations for understanding the difference in immune responses between individuals or the activation of immunological reactions to diseases.

The MHC, referred to as the HLA in humans, is a multigene family located on the short arm of chromosome 6 (band p21.3). The MHC-related glycoproteins encoded by *HLA* have a central role in regulating immune responses and controlling cell-to-cell interactions. In the mammalian genome, the MHC region is considered to be one of

the most gene-condensed regions where many essential genes relating to immunology have converged. The MHC region also displays an extremely high level of polymorphism, especially at the classical HLA loci with over 65,000 HLA and HLA-related alleles (HLA Nomenclature 2021). MHC molecules are classified into three major classes: MHC-I, II and III. While MHC-I and II alleles encode cell-surface glycoproteins involved predominantly in the antigen presentation pathway and the activation of the adaptive immune response, MHC-III encodes molecules playing crucial functions in the innate immune system and humoral immune response. In this study, the focus is on the MHC-I region.

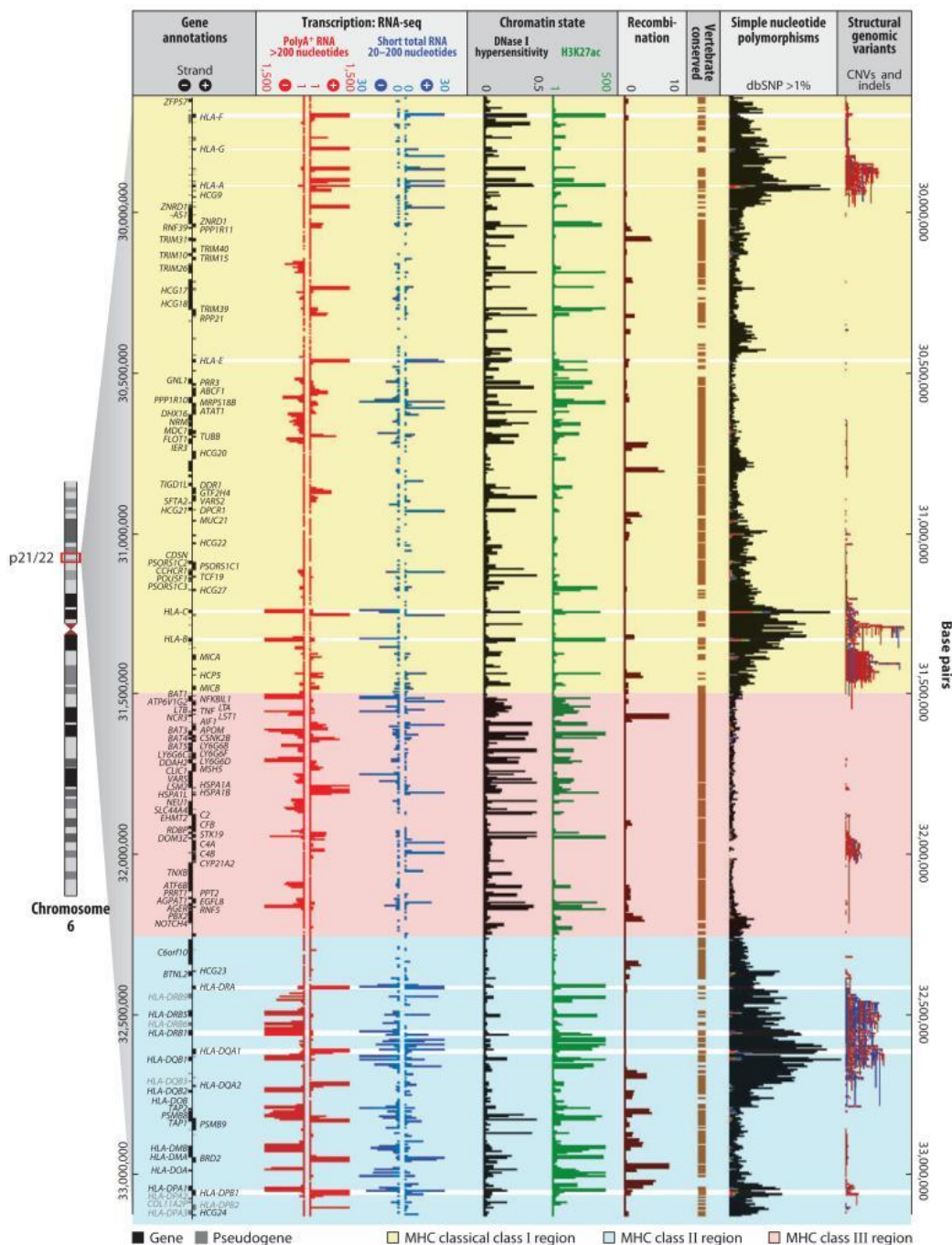


Figure 1.1 Genomic landscape of the MHC.

The MHC region is located on the short arm of chromosome 6 (base pair positions 29,640,000–33,120,000 from the Genome Reference Consortium Human Build 37, hg19), containing the MHC-I, II, and III regions. These MHC regions contains over 65,000 alleles divided into more than 39 different genes. Illustration was obtained from Trowsdale et al [5]

Since the first HLA class I molecule was isolated in 1968 by Mann et al, there are currently nearly 11000 HLA class I alleles has been announced in 2017 [6, 7]. MHC-I molecules are classified into 6 types of HLA: A, B, C, E, F and G. A, B and C are considered as the classical MHC-I molecules. MHC-I molecules are composed of two main components: a heavy chain (α chain) encoded by HLA-I and the light chain β 2-microglobulin (B2M) encoded by the *B2M* gene on human chromosome 15. The heavy chain associates non-covalently with B2M and presents endogenously derived peptides to immune cells for antigen recognition. MHC-I molecules are expressed on the membrane surface of most nucleated mammalian cells and interact with CD8⁺ cytotoxic T lymphocytes (CTL) and natural killer cells (NK). The activation of CTLs depends on the self or foreign antigens presented by MHC-I molecules, which also act as ligands for killer cell immunoglobulin-like receptors (KIR) expressed on NK cells [8].

Similarly, to MHC-I, MHC-II molecules present peptides at the plasma membrane but are only expressed by antigen-presenting cells (APCs) including dendritic cells (DC), macrophages, B cells, epithelium cells of the thymus and endothelial cells. MHC-II molecules are heterodimers composed of alpha (α) and beta (β) chains. The open binding groove formed by α and β chains accommodates extracellular protein-derived peptides of 13-25 residues in length which are presented to CD4⁺ T helper cells [9]. Therefore, MHC-II molecules mediate CD4⁺ T helper cells which can regulate antibody responses by B cells and modulate CTL activity by releasing cytokine signals [10]. MHC-II molecules are classified into three classical types which include DP, DQ and DR, and two non-classical types including DM and DO. DM molecules function in supporting peptides to be loaded onto classical MHC-II molecule within a specialized compartment within the endocytic pathway i.e., the MHC-II loading compartment

(MIIC)/multivesicular body [11, 12]. On the other hand, DO molecules are mainly expressed in B cells and are proposed to have an inhibitory effect on DM function [11, 12]. The polymorphism is not as extensive as class I but there is the added complexity of mixed isotype expression, such as DQ α pairs with DR β chains, and haplotype pairing, such as different DQ α chains pairing with different DQ β allele chains.

The MHC-III region, located between the HLA class I and II regions, contains around 60 genes involved in inflammation, immunity, and regulation of immunity responses [13]. The first example of important components located within MHC-III region are *C4*, *C2* and *BF* (factor B), which are parts of the complement system having roles in enhancing the ability of antibodies and phagocytic cells to clear pathogens [13]. Secondly, the *TNF* gene is also encoded within the MHC-III region and plays an important role as a proinflammatory cytokine in regulation of cell proliferation and differentiation as well as innate and adaptive immune responses. Mutations in some MHC-III genes can lead to diseases, for instance mutation in *CYP21* can lead to congenital adrenal hyperplasia which affects the adrenal glands [14].

1.2 MHC class I

1.2.1 Heavy chain of MHC-I

MHC-I heavy chains (HC) are encoded within the HLA-I loci – HLA-A, B, C, E and G regions of the MHC locus. MHC-I HC molecules are approximately 45 kilodalton (kDa) in size and their structure can be divided into three distinct parts: the cytoplasmic, transmembrane, and extracellular domains. Although the cytoplasmic domain of MHC-I molecules was considered to have no functional activities in the past [15], recent studies reported several functions contributed by this component. For example, the dimers of two fully folded MHC-I are formed by the disulphide bond between the cytoplasmic tail domain cysteines and found on small secretory vesicles known as exosomes [16]. The half-life of HLA-G, a non-classical MHC-I molecule expressed abundantly on extravillous trophoblasts in physiological conditions, was also found to be prolonged due to the loss of an endocytosis motif in the cytoplasmic tail of the heavy chain [17]. The cysteine at position 309, located near the interface of inner leaflet of the cytoplasm and lipid bilayer, is only found within the transmembrane region of HLA-B and -C alleles which are major ligands for NK inhibitory receptor (NKIR). In Davis's study [18], cytotoxicity assays results showed that 68% NK cells were inhibited by cells expressing HLA-Cw6 but efficiently lysed target cells expressing HLA-Cw6 with Cys³⁰⁹ mutated to tryptophan. Therefore, the transmembrane domain of HLA-B and -C molecules was proposed to play a critical physiological function in interaction with a subset of NK cells.

The extracellular region of MHC-I molecules is composed of three domains: $\alpha 1$, $\alpha 2$ and $\alpha 3$ (Figure 1.2). Amongst the three domains, $\alpha 3$ exhibits the least polymorphism and predominantly contains hydrophobic amino acids. The $\alpha 3$ domain adopts an

immunoglobulin-like fold and forms non-covalent intermolecular interactions with the B2M molecule [19]. While $\alpha 3$ residues of the heavy chain are highly conserved and contains binding sites for the CD8 co-receptor on cytotoxic T lymphocytes, $\alpha 1$ and $\alpha 2$ residues, which together form the peptide-binding groove, are much more polymorphic and will be covered in more detail in the next section.

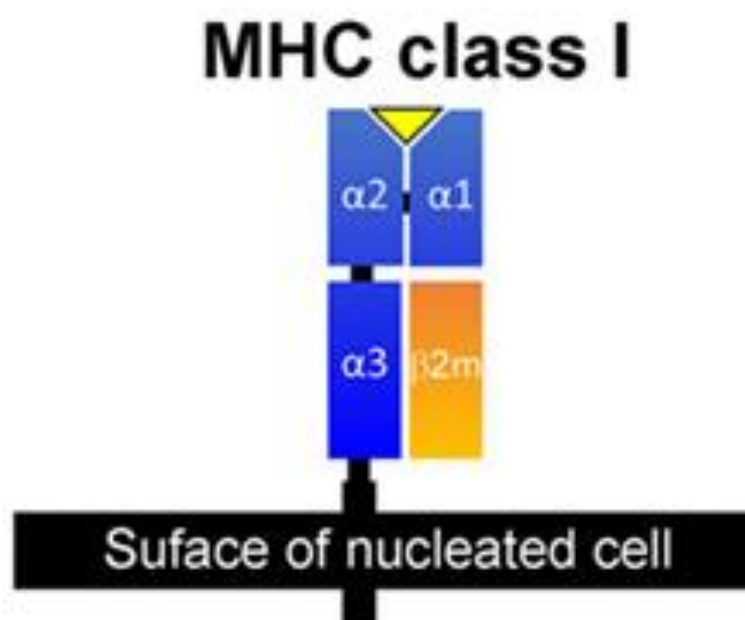


Figure 1.2. Structure of major histocompatibility complex (MHC) class I.

Domain topology of a MHC-I complex consisting of a heavy chain HLA (blue), a light chain B2M (orange), and peptide (yellow). Figure is adapted from Wieczorek et al [20].

1.2.2 Polymorphisms of MHC-I

HLA class I molecules are encoded by three classical HLA class I genes, *HLA-A*, *HLA-B* and *HLA-C* within the MHC-I region. HLA class I typing began in the 1960s and thanks to new sequencing technologies, more than 24,000 *HLA-A*, *HLA-B* and *HLA-C* alleles have been deposited in the database according to the latest version of the IPD-IMGT/HLA Database (May 2022). In Robinson's study, he illustrated that *HLA-B* and *HLA-C* alleles are more similar to each other than either is to *HLA-A* (Figure 1.3) [7]. These results support the theory that *HLA-C* alleles originated from the duplication of *HLA-B* alleles [7]. The majority of variations within MHC-I molecules are concentrated in exons 2 and 3 which encode the $\alpha 1$ and $\alpha 2$ domains. The nucleotide substitution events appear frequently at >95% of all positions in each of three *HLA* genes, therefore leading to substitutions at every single amino acid in the $\alpha 1$ and $\alpha 2$ domains of *HLA-A*, *-B* and *-C* alleles. Based on the rate of new alleles discovered, it has been estimated that there exists 2-3 million *HLA-A*, *-B* and *-C* alleles in the worldwide population [7].

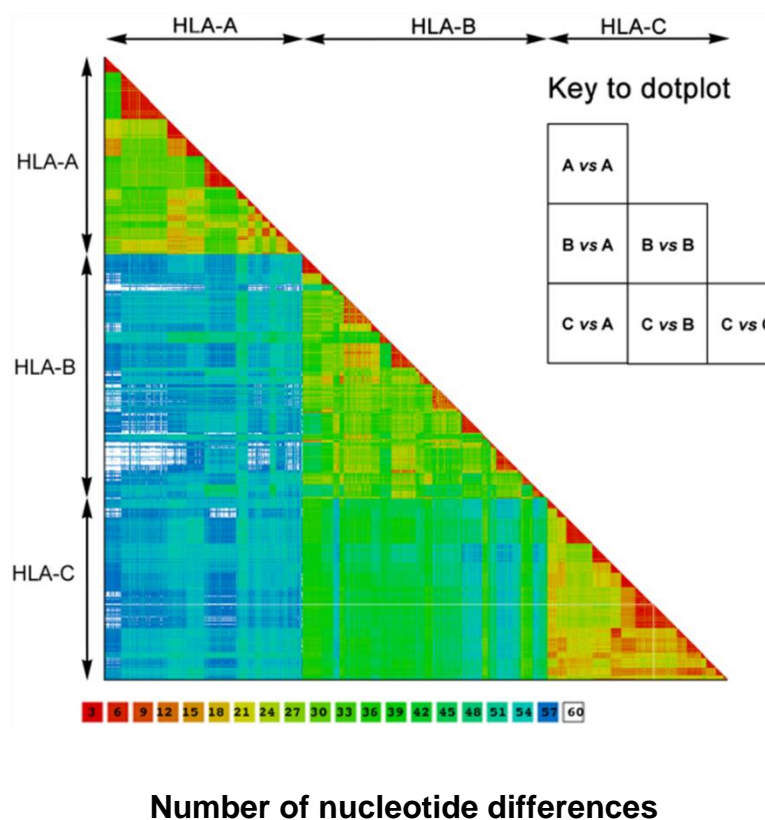


Figure 1.3. Pairwise comparison of *HLA-A*, *-B* and *-C* alleles.

A colour scale demonstrates the number of nucleotide differences in each base pair compared between *HLA-A*, *-B* and *-C* alleles. Red and blue represent the most closely and loosely related alleles, respectively. Graph was obtained from Robison et al, 2017.

While residues of the $\alpha 3$ heavy chain domain are highly conserved and contain binding sites for the CD8 co-receptor on cytotoxic T lymphocytes, $\alpha 1$ and $\alpha 2$ residues, which together form the peptide-binding groove, are much more polymorphic. Polymorphisms of the $\alpha 1$ and $\alpha 2$ domains lead to diversity in the peptide spectrum presented by these molecules. The advantage of such peptide variation with respect to presentation by MHC-I is that it increases the opportunities for the adaptive immune system to generate specific weaponry against pathogens. Furthermore, the $\alpha 1$ and $\alpha 2$ domains also contain the binding sites for lymphocytes receptors [21]. Although the

interaction between HLA-I and lymphocytes assist in the detection and removal of infected cells, MHC molecules form a substantial barrier for successful bone-marrow and organ transplantation between donors and recipients of differing HLA types. Furthermore, despite donors being MHC matched, the variation in individual peptide pools can lead to more chronic organ rejection on the basis of these 'minor' histocompatibility differences [22]. Another important clinical corollary associated with MHC polymorphism is that numerous HLA alleles and haplotypes have been demonstrated to be associated with human diseases such as Ankylosing Spondylitis (AS), Behçet disease, and Psoriasis associated with HLA-B*27, HLA-B*51, and HLA-C*06, respectively [23].

1.2.3 The Light Chain of MHC-I; Beta-2 Microglobulin

The second core component of MHC-I molecules is B2M, a small protein of 11.7 kDa. B2M is encoded by the *B2M* gene which is located on chromosome 15 and has limited genetic variations [24]. B2M is expressed by all nucleated cells and found in most biological fluids, including urine, serum and synovial fluid [25]. B2M consists of 7 β -strands, which form two β -sheets linked by a disulphide bridge and has the typical β -sandwich shape of an immunoglobulin (Ig) domain [25]. Trp60 and Trp95 are conserved tryptophans and play complementary roles in the structure of the B2M molecule. While Trp60 is critical for enhancing the association of B2M with HLA-I, Trp95 is related to the stability of B2M and preventing the formation of aggregates [24, 26]. In addition to classical MHC-I molecules, B2M also associates with other MHC-like structures including a cluster of differentiation 1 (CD1) – a foreign lipid antigen binding molecule [27], human hemochromatosis protein (HFE) which can bind to

transferrin receptor 1 (TfR1) and weakens the uptake of transferrin-bound iron in cells [28], and neonatal Fc receptor (FcRn) – an important molecule involved in the binding and transport of IgG and humoral immunity [29]. B2M also has a special role in regulating the survival, apoptosis, proliferation, and metastasis of cancer cells, which makes B2M a potential target for multiple diseases, especially in cancer therapeutics (Table 1.1). For example, in Rowley's study [30], purified human B2m stimulated cell proliferation and significantly reduced doubling time in rat PS-1 prostatic stromal cells and human PC-3 prostatic carcinoma cells. The high concentration of free-circulating B2M, as a result of renal failure, can also lead to the formation of amyloid consisting of B2m, heparin, collagen, and apolipoprotein E and enhance WT-hB2m aggregation *in vitro* [31, 32].

Table 1.1 Examples of different functions and applications of B2M.

Table was adapted and updated from Li, Dong and Wang's research, 2016.

Year	Targeting	Mechanism	Application	Ref
1992	Osteoblast	Mitogen	Therapeutic target	[33]
1993	MHC-I reconstruction	MHC-I or like molecules stabilizer	Vaccine adjuvant	[34]
1995	Antagonistic activity to transforming growth factor beta 1	Hormone/growth factor receptors	Immune regulation and cell proliferation	[30]
2001	Leukemic cell-bearing mice	Apoptosis-inducing activity via activation of caspase-3 and nuclear factor-kappa B	Therapy for leukemia	[35]
2002	Myeloma cells	Negative growth regulator, induce cell apoptosis	Therapeutic strategy	[36]
2003	Human lymphoblastic leukemia cell line	Induce apoptosis via increasing reactive oxygen species	Therapy for leukemia	[37]
2006	Prostate cancer bone metastasis	Signaling and growth promoting factor	Therapeutic target	[38]
2006	Prostate cancer bone metastasis	Signaling and growth promoting factor	Therapeutic target	[39]
2007	IL-6 and IGF-I receptors and signaling pathways	Anti-B2M mAbs redistribute or block IL-6 and IGF-I receptors or signaling pathways	Apoptosis of myeloma cells	[40]
2008	Mesenchymal stem cells	Growth stimulator	Prognostic marker and therapeutic target of cancers	[41]
2019	Joint and cartilage tissue	WT-hB2M amyloid formation	Treatment of long-term hemodialysis	[32]
2021	serum B2M levels	Atrial Fibrillation	Marker of cardiovascular disease	[42]

1.2.4 MHC-I Protein Folding and Assembly

MHC-I molecules fold and assemble within the lumen of the ER, with the assistance of lectin chaperones, members of the Protein disulfide-isomerase (PDI) family of oxidoreductases and specific accessory molecules. The chaperone mediated folding events governing MHC-I folding and assembly which can be broadly allocated into two steps: (i) early folding events which lead to the association of heavy chain and B2M and (ii) the formation of the Peptide Loading Complex (PLC) in order to acquire and optimize the peptide cargo [43].

MHC-I HC polypeptides are concomitantly synthesized and translocated into the lumen of the ER via the Sec61 translocation channel [44]. Nascent MHC-I HCs are post-translationally modified with the $\text{Glc}_3\text{Man}_5\text{-}_9\text{GlcNAc}_2$ (Glc, glucose; Man, mannose; GlcNAc, N-acetylglucosamine) oligosaccharide. This oligosaccharide moiety is linked to Asparagine at position 86 within the Asn-X-Thr/Ser (x representing any amino acid) glycosylation motif of the MHC-I HC. The alpha 1,2-glucose and alpha 1,3-glucose residues of the MHC-I linked oligosaccharide are trimmed by glucosidases I (Gls I) and II (Gls II), respectively, leading to the formation of a monoglucosylated unit, $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$, which is recognized by the transmembrane lectin-like chaperone calnexin (CNX) [45, 46]. When *Asn86* is mutated, the CNX-HC association is not abrogated completely, suggesting other protein-protein interactions also exist [47, 48]. Concomitantly, CNX opens an extended proline-rich domain (P domain) and recruits the oxidoreductases (ERp57) to promote folding events or protein disulfide isomerases (PDI) to synthesise disulphide bonds [49, 50]. After the nascent MHC-I is completely folded, Gls-II removes the terminal glucose and calnexin is displaced following the recruitment of calreticulin (CRT) [51]. At this stage, MHC-I HC

has assembled with B2M and is ready for the next stage which is PLC formation [52]. Early folded heterodimers of MHC class I/B2M are still unstable so CRT is recruited via their monoglucosylated *N*-linked glycan which helps form the PLC [53, 54]. MHC-I, B2M, ERp57 and CRT are tethered to the transporter associated with antigen processing (TAP) via the specific accessory molecule tapasin [55, 56]. TAP consists of two subunits: TAP1 and TAP2 which together are responsible for delivering preferentially ~8-16 amino acid long peptides from the cytosol into the ER for binding by the MHC-I molecule [57].

Within the ER, the MHC class I specific accessory molecule tapasin, bridges MHC-I molecules to TAP and holds MHC I molecules in a peptide-receptive state [58]. Tapasin can shape the repertoire of presented peptides by maintaining the peptide-binding process of MHC-I in a slow-off rate [59]. Tapasin selects for optimal MHC-I peptide binding by mediating the exchange of suboptimal low affinity peptides for peptides which bind more stably [60]. Zarling's study [61] showed the binding affinities of ligands eluted from tapasin-positive and negative cells were equivalent. Zarling's observations suggests tapasin may function more as a facilitator of the peptide-binding process rather than an editor of the peptide repertoire. Recently, a second MHC-I molecule specific accessory molecule that functions as a peptide exchange catalyst has been identified i.e. the transporter associated with antigen processing binding protein-related (TAPBPR) protein [62-64]. In contrast to tapasin, TAPBPR operates outside the peptide-loading complex [62] and can function to replace some peptides of low/medium affinity to enhance peptide selection and stability of MHC-I molecules [63].

Optimal peptides of MHC-I molecules are usually 8-9aa in length, expressing appropriate anchor residues. ER resident aminopeptidase, ERAP 1, can bind to peptides longer than 8-9 residues and trigger a conformational change which activates its hydrolytic activity [65]. Therefore, ERAP1 proteins are responsible for trimming longer peptides to optimal lengths [66]. Optimal peptides are required to stabilize MHC-I complexes when these molecules are not bound to chaperones within the ER and allow the MHC-I to pass the ER quality control system before transportation to the cell surface for presentation to CD8⁺ T cell [67, 68].

1.3 MHC-I specific antigens

1.3.1 Selection and presentation of MHC-I specific antigens

MHC-I specific antigens are short peptides which form the third subunit of the tripartite MHC I complex and are required to display intracellular proteins to cytotoxic T cells (CTLs). The binding groove of MHC-I complexes usually present peptides of 8-10 amino acids in length which bind within pockets by appropriate anchor residues [69]. Peptides that are longer in length are still capable of capture by MHC-I but are trimmed to the correct size by the endoplasmic reticulum aminopeptidases (ERAP) 1 and 2 [70]. Most other amino acids can fill the free spaces outside the anchor residues. Each individual MHC-I molecule can in theory present a peptidome of approximately 6×10^6 different antigens [71]. However, most peptides are non-existent or cannot be presented due to limited number of cell membrane MHC-I complexes (about 200,000) [72]. Therefore, it is estimated that there are less than 10,000 different peptides presented by any one cell [73]. Nevertheless, this array of peptides is deemed sufficient to protect against most pathogens. Importantly, the different MHC molecules between individuals can broaden the choice of peptides presented to T cells against the same pathogen and benefit the survival of the population.

Many of the key steps by which MHC-I molecules present peptide have been established through studies over the last 30 years [74]. Firstly, peptide fragments are derived from autologous normal proteins, mutated self-proteins which can be expressed by mutated genes (e.g. in cancers), foreign polymorphic genes (e.g. from graft transplants), and pathogens such as viruses and bacteria. Peptides are generally generated via the proteolytic activity of the proteasome. Most of these fragments are subsequently degraded into small peptides by cytosolic peptidases including

aminopeptidases, carboxy peptidases and ectopeptidases [75]. Peptides are transported into the ER by a peptide transporter called transporter associated with antigen processing (TAP) [74, 76]. Van Kaer's study demonstrated that MHC-I molecules expressed by TAP-deficient cells are unstable and not capable of presenting antigens to CTLs [77]. The complete TAP system is formed of a heterodimeric complex of TAP1 and TAP2 subunits which are encoded by *ABCB2* and *ABCB3* genes lying in the cluster of genes for antigen processing within the MHC-II region [5]. After binding to proteasome generated peptides, the TAP1/2 heterodimer translocate peptides from the cytosol into the ER via the hydrolysis of ATP which binds to the cytosolic nucleotide binding domains [78]. At the same time, MHC I molecule heavy chains and B2m undergo a degree of folding (as mentioned in 1.2.2) before binding to optimal peptides within the peptide loading complex (PLC) consisting of TAP, tapasin, calreticulin, PDI and ERp57 [79]. At this stage, multiple peptides are respectively "scanned" by MHC-I molecules until low off-rate peptides are bound [80].

Peptides longer than the optimal length can be trimmed by ER resident aminopeptidases including ERAP1 and ERAP2. ERAP1 and ERAP 2 consist of four globular domains and have many similar structural characteristics, exhibiting 50% amino acid sequence identity [81]. ERAP 1 shows preference for non-polar residues with a wide range of cleavage efficiency depending on the N-terminal side chain of the peptide [82]. ERAP1 peptidase preferentially cleaves all peptide bonds except those involving proline and ERAP2 can cleave efficiently especially at arginine residues [82, 83]. The target peptide spectrum of ERAP1 is 9-mers or longer but that of ERAP2 is 9-mers or shorter [84]. Non-specific peptides, which are unable to bind to any MHC-I molecules, are transported back to the cytosol for degradation [85]. The completed

MHC-I complexes containing appropriate peptide are then transported to the plasma membrane, ready to be recognized by CTLs.

While evolution promotes the polymorphism of MHC-I complexes, cancers as well as viruses co-evolve to develop mechanisms to escape detection from the immune system by evading or influencing the antigen presenting process. Pathogens can inhibit MHC molecules themselves by targeting the peptide generation and acquisition stage. For instance, human cytomegalovirus (HCMV) encodes proteins (such as US2, US3, US6 and US11) which inhibit the TAP complex or induce the degradation of MHC-I complexes [86]. Additionally, HCMV miR-US4-1 has been demonstrated to downregulate ERAP1 expression which can lead to the escape of infected cells from HCMV-specific CTLs [87]. Meanwhile, Herpes Simplex Virus (HSV) encodes ICP47 which binds to the cytosolic face of TAP1/TAP2, preventing the transportation of viral peptides into the ER by blocking the TAP translocation pore, whilst US6 encoded by HCMV binds to the ER luminal face and interferes with ATP hydrolysis which occurs via the nucleotide binding domains on the cytosolic face of the TAP complex [88].

Downregulation of MHC-I complexes has also been reported in many types of pediatric and adult tumors, often correlating with worse prognosis [89]. Irreversible MHC-I defects are caused by genetic mutations within HLA class I, *B2M* or *TAP* leading to structural alterations of heavy and light chain, or peptide availability, respectively [90-92]. For example, Maleno's study [90] reported that 35% of bladder carcinomas had loss of heterozygosity (LOH) at chromosome region 6p21.3 region leading to HLA haplotype loss. Additionally, tumors can upregulate factors like transforming growth factor β (TGF- β) which can impair the function of NK cells and prevent their infiltration into the tumour region where cancer cells express low levels MHC-I [93].

Downregulation of MHC-I molecules can be the result of cancer-induced defects affecting epigenetic and post-transcriptional factors such as nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B), Interferon regulatory factors (IRFs), and NLR Family CARD Domain Containing 5 (NLRC5) [89]. Therefore, understanding tumor and pathogen escape mechanisms from antigen presentation could help develop methodology or therapies to either enhance MHC-I molecules presenting desired peptides for the activation of effective T cell immune responses and/or preventing their expression from being downregulated.

1.3.2 Manipulation of MHC-I peptide for clinical application

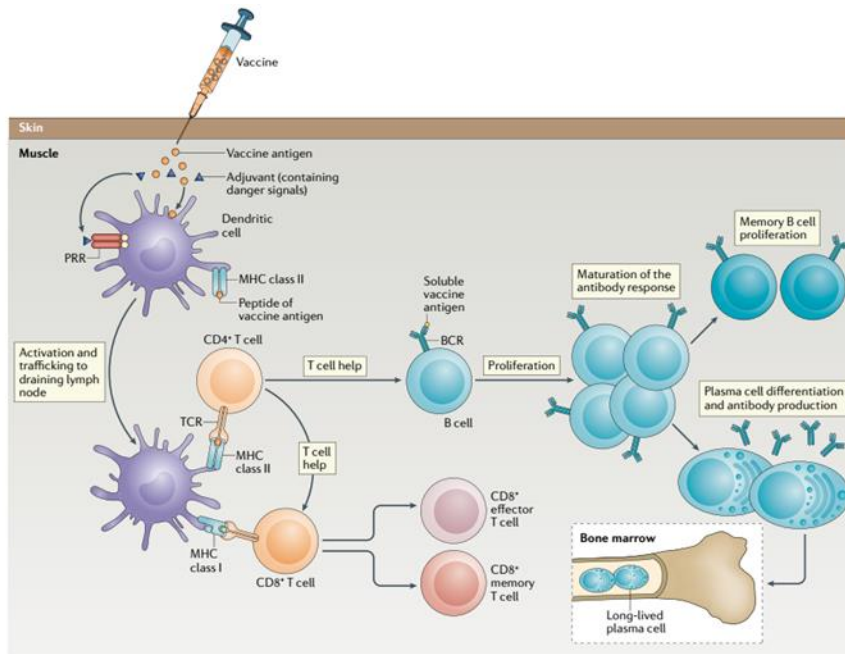
With a key role in the regulation of the immune response, manipulation of antigens presented by MHC-I molecules has become a potential approach in disease prevention and treatment. Indeed, manipulation of MHC-I peptides and their presentation have been used in the development of vaccine technologies. More recently, dendritic cell vaccines have been improved for cancer immunotherapy and the MHC single chain trimer (SCT) technology was developed to directly present desired peptides for immunogenicity.

1.3.2.1 Vaccines

Since the first vaccine was described by Edward Jenner against smallpox in the 18th century, the history of human health care has witnessed a huge jump in the development of vaccines against other lethal diseases including mumps, measles, rubella, and varicella. A vaccine is a biological product containing pathogen-derived antigens which induce immune responses to provide protection (Figure 1.4A). After delivering the vaccine containing antigens and adjuvant via oral, intranasal, subcutaneous, or intramuscular routes, pathogenic antigens are uptaken and

presented on MHC-I complexes of nucleated cells and MHC-II by dendritic cells (DCs). Under the interaction of adjuvant and pattern recognition receptors (PRRs), danger signals are released and induce DCs to translocate to the lymph node where they present peptides to T cells. Subsequently, CD8+ T cell are selected and expanded to become CD8+ effector T cells which can recognize and eliminate the cells presenting pathogenic antigens. On the other hand, CD4+ T cells can help drive B cell differentiation into plasma cells which produce antigen-specific antibody. Minor populations of CD8+ and CD4+ effector T cells and B cells differentiate into memory T and B cells, respectively, which generate rapid immune responses on the next encounter against pathogens.

A



B

Type of vaccine		Licensed vaccines using this technology	First introduced
Live attenuated (weakened or inactivated)		Measles, mumps, rubella, yellow fever, influenza, oral polio, typhoid, Japanese encephalitis, rotavirus, BCG, varicella zoster	1798 (smallpox)
Killed whole organism		Whole-cell pertussis, polio, influenza, Japanese encephalitis, hepatitis A, rabies	1896 (typhoid)
Toxoid		Diphtheria, tetanus	1923 (diphtheria)
Subunit (purified protein, recombinant protein, polysaccharide, peptide)		Pertussis, influenza, hepatitis B, meningococcal, pneumococcal, typhoid, hepatitis A	1970 (anthrax)
Virus-like particle		Human papillomavirus	1986 (hepatitis B)
Outer membrane vesicle	 Pathogen antigen Gram-negative bacterial outer membrane Polysaccharide	Group B meningococcal	1987 (group B meningococcal)
Protein-polysaccharide conjugate	 Carrier protein	<i>Haemophilus influenzae</i> type B, pneumococcal, meningococcal, typhoid	1987 (<i>H. influenzae</i> type b)
Viral vectored	 Viral vector Pathogen gene Viral vector genes	Ebola	2019 (Ebola)
Nucleic acid vaccine	 DNA RNA Lipid coat	SARS-CoV-2	2020 (SARS-CoV-2)
Bacterial vectored	 Pathogen gene Bacterial vector	Experimental	-
Antigen-presenting cell	 Pathogen antigen MHC	Experimental	-

Figure 1.4. Different types of vaccines and their various immunogenicity mechanisms.

(A) The generation of immune responses to vaccination. Vaccine antigens are delivered inside the body and are taken up by DCs. These antigens are then presented on MHC-I and II molecules and along with the expression of co-stimulatory molecules activate T cells with peptide specific TCRs. In this way, vaccine technology directs the immune system to be activated in the desired direction. (B) Classification of vaccines. Different types of vaccine developed against pathogens. Graphic illustration was obtained and modified from Pollard and Bijeker [94].

Vaccines are classified into two major groups, (i) live or (ii) non-live, (Figure 1.4B). Live vaccines are generally attenuated pathogenic organisms which are still capable of replicating to induce responses but avoid symptomatic disease. Although live vaccines provide a strong immune response, the major drawback is the risk of causing significant disease. Non-live vaccines contain only pathogenic components or killed organisms. Non-live vaccines are normally co-injected with an adjuvant to enhance their ability to generate immunogenicity. The discovery of numerous microorganisms, evolution of cell culture and the introduction of protein recombination and whole-genome sequencing platforms has promoted outstanding development of vaccine technology. During the Covid-19 pandemic since 2020 until now, many vaccines have been developed and approved in record time, accompanied by first-time technologies such as mRNA vaccines such as those developed by Pfizer-BioNTech and Moderna for COVID-19. Although vaccines provide excellent prevention against infectious disease, this technology faces many challenges and limitations in cancer treatment. One of the reasons is that the antigen array of tumors may not be specific to the tumor alone but is also shared by healthy tissues which can lead to the potential risk of developing autoimmunity. Nevertheless, many other vaccine platforms, such as

antigen-presenting cells, T cell and bacterial vectors, are continuously being developed and tested to broaden the arsenal against diseases [95].

1.3.2.2 *Engineering dendritic cell (DC) vaccines*

Dendritic cells (DCs) are professional antigen presenting cells which are effective at taking up and presenting antigens to effectively activate T cells, especially naïve T cells due to the expression of co-stimulatory molecules [96]. In DC therapy, hematopoietic stem and progenitor cells (HSPC) or monocytes, which contain DCs, are isolated from peripheral blood and treated with different cytokines to induce differentiation and induce maturation. Following differentiation, TAAs are loaded on the DC's MHC-I molecules via two different approaches: (i) the secretory pathway of antigen presentation and (ii) the vacuolar pathway or cross presentation pathway, in which TAAs can be trimmed by lysosomal proteases and directly loaded onto MHC-I in the endosome. Various preclinical and clinical studies have demonstrated the immunogenicity and safety of DC therapy[97]. However, the results of clinical responses has been limited due to multiple reasons including low migratory capacity, cytokine release and especially insufficient antigen presentation [98]. Many solutions have been applied to enhance TAA presentation, including expression of an extracellular vesicle-internalizing receptor (EVIR) to increase antigen uptake, or disruption of YTH N6-Methyladenosine RNA Binding Protein 1 (YTHDF1) which reduces translation of lysosomal cathepsins to decrease TAA degradation [96, 99]. Recently, induced DCs (iDCs) have been successfully differentiated from Induced Pluripotent Stem Cells (iPSCs) which are embryonic-like pluripotent cells programmed back from skin or blood cells [100].

1.3.2.3 *MHC Single Chain Trimer (SCT) technology*

The polymorphism of HLA types and diversity of peptides result in many different MHC-I molecules being expressed between individual cell types and cells from individuals. This flexibility in the composition of MHC-I molecules also enhances the possibility of pathogen-specific peptides being presented to generate effective CD8+ T cell responses. However, both virus and cancer cells has developed their own mechanisms to evade or escape from antigen presentation. Therefore, scientists have paid their attention to develop methods for manipulating MHC-I molecules to improve CD8+ T cell responses and expand CTLs populations for therapeutic purposes [101]. One of these methods is the development of MHC-I single-chain trimers (SCT), which was first described by Hansen and his colleagues in 2002 [102]. SCTs has been intensively studied and have been specifically evaluated as potential DNA vaccine platforms (Table 1.2). However, the SCT approach encountered certain limitations which has led this technology to gradually receive less attention.

Table 1.2. Summary of studies using MHC-I single-chain trimers to induce CTL responses.

Year	MHC type	Antigens/peptide sequence	Origin of peptide	Ref
2002	H-2K ^b	OVA (SIINFEKL)	Ovalbumin	[102]
2005	H-2K ^b	E6 antigen	HPV	[182]
2005	HLA-E	Beta2m (VMAPRTLIL)	Human	[183]
2007	HLA-A2	N220 (LALLLLDRL)	SARS-CoV	[184]
2007	H-2K ^b	E6 antigen	HPV	[185]
2007	HLA-E	VMAPRTLVL	CMV	[186]
2007	HLA-A2	Mesothelin (aa540-549)	Human cancer	[187]
2009	H-2D(b)	ASNENMDAM	Influenza	[188]
2010	HLA-A2	SVG9	West Nile Virus	[189]
2010	H2-Kb	TRP-2 (SVYDFVWL)	Melanoma	[190]
2011	HLA-A2	SLYNTVATL	HIV	[191]
2014	HLA-A2	Matrix (GILGFVFTL)	Influenza A	[192]
2015	H-2D ^b	VEGFR2 (aa400-408)	Human cancer	[193]

MHC-I SCT molecules are a fusion between all components in the order from N- to C-terminal as follows: specific peptide, linker 1, B2M, linker 2 and HLA class I heavy chain (Figure 1.4B) [102]. SCTs are often constructed in the form of a DNA plasmid and delivered into cells for protein expression. After transfection, SCT DNA is expected to be transcribed in the nucleus and translated within the ER. Via the MHC-I assembly pathway, naked SCT residues are folded and potentially bypass all quality control checking before translocating to the cell surface. Compared to conventional

MHC-I molecules, SCTs are assembled independently of TAP as demonstrated by their expression in TAP-deficient cell lines [191]. SCT expression also requires a match between HLA class I types and their HLA-specific peptides. A study of SCT transfection into TAP-deficient cells showed that an HLA-A2 SCT encoding an HLA-A2-restricted HIV epitope (SLYNTVATL) exhibited enhanced expression at the plasma membrane than an HLA-A2 SCT encoding a HLA-B27-restricted influenza virus epitope (SRYWAIRTR) [101]. Studies have also described that specific T cells can recognize SCT complexes and induce immune responses without any cross-reactivity caused by the presence of linkers between the presented peptide, B2M and heavy chain [194, 195]. NK cell surface receptors can also recognize MHC-I SCT constructs and have been used to inhibit NK cells from being activated and thus eliminating cells expressing the SCT [196].

Due to the unique properties of MHC-class I SCTs, they have significant advantages in generating CTL responses but also possess difficult obstacles to overcome before application for clinical use. Various studies using MHC-I SCTs as DNA vaccines provided very optimistic results in generating as well as detecting CD8+ T cell responses (Table 1.2). As mentioned above, many viruses and tumor cells can inhibit the presentation of pathogenic antigens, but SCT technology brings the potential solution to overcome some of these evasion mechanisms. For instance, the endogenous synthesis of SCTs is independent of the peptide transporter associated with antigen processing, therefore, SCTs can still induce CTL responses against cells infected by TAP-targeting viruses such as Herpes simplex virus (HSV). Additionally, SCT technology can be used to generate artificial antigen-presenting cells (aAPCs) which present highly specific antigens to activate and expand CD8+ effector

lymphocytes [197]. In Obermann's study, aAPCs were developed by transfection of MHC-I negative Daudi cells (which lacks expression of B2m) with an SCT construct ensuring presentation of the peptide influenza virus matrix M1₅₈₋₆₆ (GILGFVFTL) - HLA-A2 complex. These aAPCs were equally efficient in stimulating influenza M1-specific T cells compared to autologous monocyte-derived DCs [197].

Despite their great potency, SCT technology still faces barriers with reduced research interest. The most significant disadvantage of SCTs is that they require delivery in the form of DNA plasmid platform. Although transient transfection of DNA platforms does not require integration of nucleic acids into the host cell genome [198], expression of SCTs will eventually be lost as host cells replicate and therefore lose the ability to repeatedly stimulate antigen-specific T cells over time [199]. Therefore, new SCT platforms are necessary to fully exploit the potency of this technology for use in different approaches such as stimulating anti-viral responses and potential immunotherapy of cancer.

1.4 Ankylosing Spondylitis and *HLA-B27* gene

1.4.1 Clinical features of Ankylosing Spondylitis

Ankylosing Spondylitis (AS) is a major subtype of the Spondyloarthritis (SpA) family of inflammatory arthritic diseases. Subtypes of SpA have common clinical symptoms and genetic predisposition. Besides AS, the other subtypes of SpA are Reactive Arthritis (ReA), Psoriatic Spondylarthritis (PsA) and undifferentiated spondylarthritis [103]. Among them, AS is the most common with the estimated prevalence of up to 1.2% [104]. AS is an autoimmune disease which seems to develop through complex interactions between environmental factors and genetic background. In AS, inflammation appears initially in sacroiliac and spinal joints, as well as their adjacent soft tissues including ligaments and tendons. In severe cases, inflammation can lead to calcification and fibrosis which later causes loss of spinal mobility and fusion of the spine, leading to the clinical manifestation known as “bamboo spine” (Figure 1.5) [105].

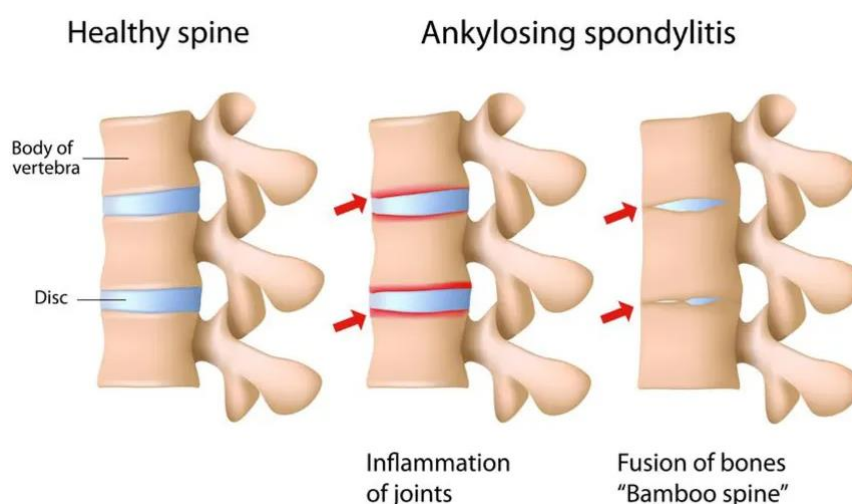


Figure 1.5. Spinal joint of healthy and Ankylosing Spondylitis individuals.

In AS cases, chronic inflammation can lead to the fusion of the spinal vertebrae and joints. Inflammation is thought to occur at the entheses and cause enteritis which is also common in AS patients Image was obtained from Christopher Shaffrey [105].

Effects of AS can appear within different parts of the body and organ systems as well as differ between patients. Ankylosing means “stiffening” and Spondylitis means “spine inflammation”, therefore, the main clinical symptom of AS is progressive spinal rigidity, causing frequent stiffness and pain in the lower back [106]. The spinal pains and stiffness often appear in the morning and after rest, but these symptoms can be relieved and improved with a warm shower or light exercise [107]. Overtime, inflammation of the spine can lead to the formation of syndesmophytes which are bony outgrowths originating from spinal ligaments and can initiate vertebrae fusion, probably as a result of an imbalance between osteoproliferation and osteodestruction. Bamboo vertebrae decrease flexibility of the spine, leading to difficulty in mobility and posture changes. As the ribs are attached to vertebrae, fusion of the spine can cause difficulty in expanding the chest and affect inhalation function. The chronic inflammation primarily develops in the sacroiliac joint, but over time it can spread to other joints including jaw, neck, ribs, shoulder blades, knees, hip, ankles and feet [108]. On the other hand, AS manifestations also include peripheral arthritis which mainly affects the lower limbs and is recorded in around 20% of AS patients [109]. Additionally, Brophy’s study has shown the interrelationship between AS and inflammation of other organs such as the gut, eyes and skin [110].

A study by Dean *et al* reported that AS prevalence estimates can vary between populations of different continents and between genders [111]. North American and European countries have the highest prevalence of AS with means of 31.9 and 23.8

per 10,000, respectively. However, there appeared to be lower prevalence within Asian, Latin American and African populations with rates of 16.7, 10.2 and 7.4 per 10,000, respectively. The estimated total amount of AS cases ranges from 4.63 to 4.98 million in Asia and from 1.3 to 1.56 million in Europe. In addition, males reportedly account for the vast majority of AS cases with the mean gender ratio being 3.4:1 (males:females).

AS symptoms often develop slowly and are predominantly diagnosed via imaging tests using X-ray and MRI to monitor changes in joints and bones [240]. Most of the current AS treatments are only capable of relieving symptoms but not reversing the disease. Anti-inflammatory treatments include anti-TNF- α monoclonal antibody (secukinumab), anti-IL-17 monoclonal antibody (ixekizumab) and corticosteroids but these treatments also lead to several side effects caused by depressing the immune system against other pathogenic factors [108]. Overall, a lack of early diagnosis and disease-modifying therapeutics of AS come from the limitation in understanding the pathogenesis of the SpAs which may involve heredity, immunity and other factors.

1.4.2 Pathogenesis of AS

Despite the intensive study during the last five decades, the mystery regarding the mechanism of AS development has not been revealed yet. In 1973, HLA-B27 has been identified to be strongly associated with SpA by two independent groups [112, 113]. However, recent genome-wide association studies (GWAS) have identified more than 100 other common loci which potentially contribute to disease risk and is dominated by HLA-B27 [114]. Therefore, the pathogenesis of AS can be categorised

into two groups: HLA-B27 and non-HLA-B27 including other MHC alleles and IL-23/IL-17.

4.1.2.1 HLA-B27 association

Despite an unclear disease mechanism, studies have demonstrated a role for HLA-B27 in the development of AS [115]. One of the key studies which illustrated a direct role for HLA-B27 was the expression of HLA-B27 as a transgene in rats. Hammer's study [116] revealed that transgenic rats expressing more than 50 copies of HLA-B27 develop spontaneous AS-like disease, with more prevalent colitis or inflammatory bowel conditions than axial skeletal disease. On the other hand, transgenic rats with less than 50 copies of HLA-B27 remain healthy. Interestingly, subsequent studies observed that transgenic rats expressing both HLA-B27 and hB2M proteins develop similar colitis conditions, but more severe arthritis and high incidence of ankylosis of the axial skeleton [117]. Furthermore, engraftment of bone marrow cells from HLA-B27 transgenic rat can induce the development of AS-like disease in previously healthy rats [118]. Thus, transgenic rat models emphasize the influence of HLA-B27 in the pathogenesis of AS as well as other pathologies associated with AS.

The prevalence of HLA-B27 is 3-8% of the world population but AS only develops in 1-2% of the HLA-B27 positive population [119]. However, HLA-B27 is found within the majority of AS cases (90-95%), and the percentage of HLA-B27 positive monozygotic and dizygotic twins developing AS are 63% and 23%, respectively [120]. Additionally, the first-, second-, and third-degree relatives of AS patients have a remarkable high risk of developing AS with 94%, 25% and 4%, respectively [121].

HLA-B27 belongs to the family of MHC-I surface proteins encoded by the HLA-B locus on chromosome 6. HLA-B27 exhibits one of the highest genetic polymorphisms with 160 known subtypes from HLA-B*27:01 to HLA-B*27:106 with different prevalence rates between different ethnicities [122]. Studies demonstrated that HLA-B*27:05, 04 and 02 are the most prevalent subtypes in AS for Caucasian, Chinese and Mediterranean populations, respectively [122-124]. By contrast, HLA-B*27:06 and :09 are reported to not be associated with AS [125, 126]. To date, the role of HLA in AS has not been elucidated. There are 4 main hypotheses proposed to explain the relationship between HLA-B27 and AS i.e. the arthritogenic peptide, molecular mimicry, misfolding and cell-surface HLA-B27 homodimer hypotheses.

Both the arthritogenic peptide and molecular mimicry hypotheses are based on the presentation of HLA-B27-specific peptides that can stimulate CD8+ T cell responses. The arthritogenic hypothesis suggests that HLA-B27 presents one or more autoantigenic peptides [127]. These self-antigens are then thought to trigger HLA-B27 restricted CTL responses against any cells presenting these peptides. For instance, collagen-derived peptides has been observed to stimulate CD8+ T cells responses in the synovial fluid of AS patients and these responses were blocked when samples were treated with an anti-HLA-B27 antibody [128]. Recently, by comparing the spectrum of peptides presented by eight of the most frequent HLA-B27 allotypes (from HLA-B*27:02 to HLA-B*27:09), 26 arthritogenic peptides had been identified to be presented in low abundance by non-disease-associated subtypes HLA-B*27:06 and HLA-B*27:09 than disease-associated HLA-B27 subtypes [129]. Further analysis revealed that only 10 peptides did not contain obvious features which could imply an

impaired binding to HLA-B*27:06 or :09. These peptides might be considered as the potential arthritogenic self-peptide candidates for further studies.

The molecular mimicry hypothesis also suggests that AS is initiated by peptides presented by HLA-B27, but of pathogenic origin. After infection, the pathogen derived peptides are thought to activate CTLs that detect and eliminate pathogen infected cells. However, these antigen(s) may be partially or completely identical to HLA-B27-derived peptides or self-peptides presented by HLA-B27 itself [130]. One of the most highlighted bacteria related to AS is *Klebsiella pneumonia* (*K. pneumonia*), which is a bacterium often found as part of the normal gut flora. On occasions, *K. pneumonia* becomes an opportunistic pathogen and has been proposed to initiate or participate in the development of AS as a triggering and/or perpetuating factor [131]. Some components of *K. pneumonia* share homologous amino acid sequences with possible self-peptides, which can potentially lead to cross reactivity between bacterial antigens and HLA peptide complexes. This cross-reactivity can subsequently induce production of anti-bacterial antibodies binding to HLA-I heavy chains on fibroblast, chondrocytes and immunocytes [132]. Another pathogen candidate proposed to lead to molecular mimicry is *Chlamydia trachomatis* (*C. trachomatis*), which is an intracellular bacteria and has been reported to elicit HLA-B27-restricted T-cell responses in Reactive Arthritis patients [133]. Beside bacterial pathogens, molecular mimicry of HLA-B27-bound self-antigens has also been suggested to originate from viruses. Based on computational analysis, several viral proteins from common viruses such as EBV, HCV or hCMV have homologous amino acid sequences with myelin basic protein [134], human transglutaminase protein [135] and GAD65 [136], respectively and are thought to participate in the development of autoimmune diseases such Multiple sclerosis. So

far, the molecular mimicry hypothesis has yet to be elucidated because no study has identified any HLA-B27 associated self-peptide resembling pathogenic antigens.

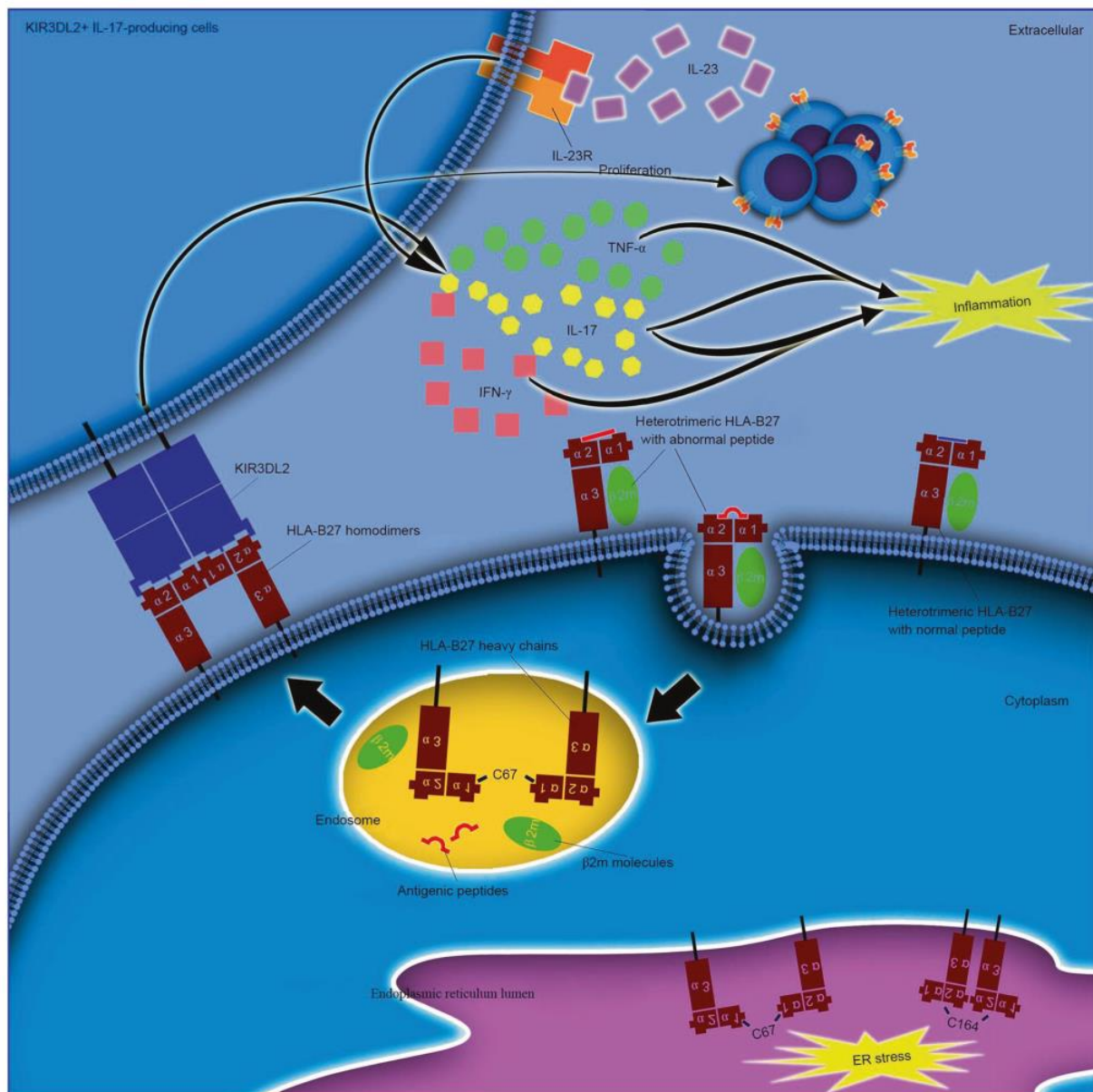


Figure 1.6. Proposed HLA-B27-related-pathogenic mechanisms of Ankylosing Spondylitis.

Properly folded HLA-B27 heavy chain associates with the B2M protein and peptide, leading to CD8+ T cell activation. However, HLA-B27 is also proposed to present auto-peptide(s) or pathogen-like peptide(s) on some occasions which can activate T cells to attack healthy cells. HLA-B27 dimers can be found within the ER and cell surface. ER dimers which are the products of inappropriate folding, might lead to ER stress and

activate the unfolded protein response. On the other hand, surface dimers, formed by the recycled heavy chain after releasing B2M and peptide, can interact with receptors, such as KIR and LILR. Subsequently, these interactions can activate KIR3DL2⁺ CD4⁺ T lymphocytes and amplify production of IL-17 and TNF- α . These cytokines are suggested to stimulate inflammatory reactions which affect the regulation of bone formation and destruction. Illustration was obtained from Zhu et al, [108].

As mentioned in section 1.2.2, the MHC-I complex assembles within the lumen of the ER under the stringent regulation of ER chaperones and the peptide loading complex prior to translocation to the plasma cell membrane. A normal MHC-I complex consists of only one heavy chain, B2M and antigenic peptide. However, in the case of HLA-B27, studies have reported the existence of three different forms of HLA-B27 heavy chains including: cell-surface heterotrimers, heavy-chain dimers and heavy-chain monomer (Figure 1.6) [137]. The abnormal forms of HLA-B27 heavy chain have been found at higher rates by AS-related MHC -I subtypes, including HLA-B*27:05, 04 and 02, compared to non-disease-associated subtypes HLA-B*27:06 and 09. HLA-B27 heavy chains have been suggested to misfold within the ER during assembly when disulfide bonds are being formed and before acquisition of B2M and peptide. MHC-I molecules contain four conserved cysteines (C) which form intramolecular disulphide bonds between C101-C164 and C203-C258 within the α 2 and α 3 domains, respectively. The HLA-B27 heavy chain possesses three additional unpaired cysteine residues at p67 (extracellular region), p308 and p325 (intracellular region). It has been proposed that the expression of C67 can lead to HLA-B27 misfolding and contribute to slower maturation kinetics than other HLA class I alleles. As a result, the aberrant HLA-B27 molecules form interchain disulphide-linked homodimers either at the plasma cell membrane, or even accumulate within the ER [138]. These aberrant features of HLA-B27 have been the foundation of the hypotheses proposing that

misfolding protein and aberrant HLA-B27 forms are involved in the disease mechanism of AS.

Under the pressure of HLA-B27 misfolding, ER function can possibly be perturbed and lead to the activation of ER stress responses by triggering the Unfolded Protein Response (UPR) (Figure 1.6) [139]. The UPR is an adaptive pathway that decreases the accumulation of misfolded protein to maintain cell function and viability. The UPR is initiated via three UPR stress sensors including activating transcription factor 6 (ATF6), protein kinase RNA-like ER kinase (PERK) and inositol-requiring protein 1 (IRE1) [140]. The association of the ER luminal domain of these UPR stress sensors with BiP maintains them in an inactive state [141]. During ER stress, it is thought that BiP is sequestered away from these effectors by the unfolded proteins thus leading to the activation of these effector proteins via autophosphorylation and oligomerization. Subsequently, the UPR stress sensors initiate a transcriptional program to boost expression of chaperones which improve protein folding as well as inhibit protein translation to decrease the misfolding protein load within the ER [142, 143]. The evidence for HLA-B27-induced UPR was reported through the upregulation of the UPR transcription factors X-box binding protein 1 (XBP1), activating transcription factor 4 (ATF4) and activating transcription factor 6 alpha (ATF6 α) as well as downstream target genes including CHOP in cells from HLA-B*27:02/B2M transgenic rats and patients with SpA [144, 145]. The UPR also intersects innate immune responses to activate upregulation of IL-23 and IFN- β with enhanced expression of cytokines augmented by interaction with toll-like receptor (TLR) agonists [146, 147].

A further hypothesis stipulates that aberrant HLA-B27 forms, such as free heavy chain (FHC) homodimers and monomers can affect the immune response and

proinflammatory responses [148]. *In vitro* studies have demonstrated that tapasin-deficient cells express higher levels of FHC homodimer than those expressing tapasin [149]. Additionally, HLA-B27 homodimers are mostly found in the synovium and gut of AS patients [108]. HLA-B27 homodimers were described to interact with killer cell immunoglobulin-like receptors 1 and 2 (KIR3DL1 and KIR3DL2) and Immunoglobulin-like transcript 4 (ILT4) [150]. Interaction between HLA-B27 homodimers and KIR3DL2 can enhance the proliferation of NK cells and stimulate Th17 CD4⁺ T lymphocytes to produce IL-17 [151]. IL-17 plays a central role in innate immune and stromal cells to promote antimicrobial functions and inflammation. Additionally, IL-17 is likely to contribute to pathogenesis of autoimmune disease involving bone immunopathology via its potential pro-osteoclastogenic effects [152]. KIR3DL2-HLA-B27 interactions were also described to enhance the survival and differentiation of KIR3DL2⁺CD4⁺ T cells which release more cytokines including TNF- α , INF- γ and IL-17 than KIR3DL2⁻CD4⁺ T cells [153]. ILT4, also known as LILRB2, is an immune receptor expressed on NK, T, B and dendritic cells. B27 homodimers and FHCs have been shown to bind ILT4 with stronger avidity than B27 heterotrimers and this interaction has been suggested to play a role in AS pathogenesis [154]. Due to the lack of antibodies which are specific for HLA-B27 homodimers and monomers, studies were limited in being able to distinguish between different types of aberrant HLA-B27 molecules.

4.1.2.2 Non-HLA-B27 association

Despite being the most strongly associated with AS, HLA-B27 only contributes 20% of disease heritability [155]. Therefore, other genetic factors also need to be considered. Beside HLA-B27, other non-HLA-B27 MHC alleles have been found to be

associated with AS in different populations. For instance, three of eight HLA-B27 negative cases were positive for HLA-B39 in Yamaguchi's study of Japanese patients [156]. HLA-B60 and HLA-B61 have been reported to increase the AS risk in Caucasian, Taiwanese and Chinese populations [157]. The above HLA class I alleles share some common features with HLA-B27 such as C67 in the case of HLA-B39. Non-HLA-B27 MHC alleles associated with AS not only included other MHC-I loci but also MHC-II alleles, for example HLA-DRB1, HLA-DQA1, HLA-DPB1, and others [119]. These MHC-II molecules may take part in AS development via their participation in antigen presentation and inflammatory processes.

Besides HLA-B27 and other non-HLA-B27 MHC alleles, GWAS studies have also reported that non-MHC variants also contribute ~7% to the pathogenesis of AS [114, 158]. Perhaps the most significant non-MHC variants associated with AS are those of *ERAP1* coding for endoplasmic reticulum aminopeptidase 1 and *ERAP2* coding for endoplasmic reticulum aminopeptidase 2 [159]. Both ERAP1 and ERAP2 have roles in optimizing HLA-I-antigens by trimming peptides to their optimal length of 9-10 amino acids within the ER [160]. While ERAP1 was reported to be associated with HLA-B27^{negative} and HLA-B40^{positive} AS cases, ERAP2 has been reported to be associated with both HLA-B27^{positive} and HLA-B40^{positive} AS cases [161]. AS-associated variants of *ERAP1* and *ERAP2* has been described to affect the amount of relevant peptide loading onto MHC-I molecules, reduce MHC-I complex maturation rates and induce ER stress as well as AS development [119]. The *ERAP1* gene is polymorphic and various polymorphisms of ERAP1 are associated with protection, disease and being neutral to disease development [162]. Besides trimming of peptide precursors, ERAP1 was also suggested to be involved in shedding cytokine receptors such as decoy IL-1

receptor (IL-1RII), TNFR1 and IL-6R α [163]. However, there is no strong *in vivo* evidence supporting cytokine receptor cleavage by different ERAP1 polymorphisms associated in AS susceptibility [164].

GWAS studies have also indicated that genetic components of the IL-23/IL-17 axis are associated with AS, with an IL-23 receptor (IL-23R) single nucleotide polymorphism (SNP) being identified to be associated with AS especially in HLA-B27 negative AS groups [165]. IL-23, a key inflammatory cytokine, acts as a bridge between innate and adaptive immunity [166], and plays an important role in maintaining and expanding T helper type 17 cells (Th17). Following triggering by IL-23, TGF- β and IL-1 β , activated Th17 cells release their effector cytokines including IL-17, IL-21, IL-22 and GM-CSF to mediate immune responses against pathogens such as fungi and bacteria [167]. In Heiner's study [168], IL-17⁺ cells were present in AS facet joints and AS patients had higher serum levels of IL-17 and IL-23, which are mainly secreted by activated DCs and macrophages in peripheral tissues such as the intestinal mucosa, joints, skin or lungs [169]. This study suggests a possible role for the innate immune system in the inflammatory processes in AS. Dysfunction of the IL-23/IL-17 axis was also reported to be associated with human immunological disorders including SpA, psoriasis, rheumatoid arthritis and IBD [170]. An uncommon coding variant sr11209026 (c.1142G>A, p.Arg381Gln) within the IL-23R cytokine molecule was considered to limit IL-23R function and subsequently affect production of IL-17 [171]. On the other hand, HLA-B27 homodimers were reported to stimulate the survival, proliferation and IL-17 production of KIR3DL2⁺ CD4⁺ T cells [149]. Additionally, several KIR3DL2⁺ Th17 cells were found to be 70% higher in SpA patients compared with a healthy B27 negative control group. Thus, both activation of the IL-23/IL-17 axis and expression of aberrant

forms of HLA-B27 may contribute to the pathogenesis of AS and other related diseases.

1.4.3 Key structural features of HLA-B27 molecules

Although several studies have been conducted to determine the role of HLA-B27 in AS, the key feature of this molecule leading to disease and the mechanism remain unclear. Current studies have focused heavily on the subtype HLA-B*27:05 as it is the most prevalent subtype in Caucasian AS patients. Besides assessing the structural factors affecting folding, one of the ways to find the key points leading to AS is to compare the structural features between disease associated HLA-B27 subtypes (HLA-B*27:02, 03, 04, 05, 07 and 08) and non-disease associated HLA-B27 subtypes (HLA-B*27:06 and 09). Comparison of the major disease associated subtypes and the 06 and 09 subtypes demonstrated there are minimal differences in amino acid sequence between the subtypes (Table 1.3).

Table 1.3. Amino acid differences between 6 of the most frequent HLA-B27 allotypes and non-AS associated HLA-B*27:06 and 09 subtypes.

Table was adapted from Sian et al, [172]

		Amino acid position																		Ref
		47	59	60	67	77	80	81	82	83	97	113	114	116	131	152	164	211		
Associated	B*27:05	P	Y	W	C	D	T	L	L	R	N	Y	H	D	S	V	C	A	[173]	
	B*27:04					S										E		G	[126]	
	B*27:02					N	I	A											[173]	
	B*27:03		H																[126]	
	B*27:07										S	H	N	Y	R				[174]	
	B*27:08					S	N		R	G									[173]	
Non-associated	B*27:06					S							D	Y		E		G	[126]	
	B*27:09													H					[125]	

An unpaired cysteine at position(p) 67 (C67) is expressed by almost all subtypes of HLA-B27, except for two rare subtypes HLA-B*27:18 (S67) and -B*27:23 (F67) [175]. Within HLA-A and HLA-B alleles, the high polymorphism of p67 indicates that this amino acid is relevant for peptide binding. Indeed, the C67 residue of HLA-B27 forms part of the B pocket of the antigen binding groove and helps determine part of the specific anchor motifs of HLA-B27 binding peptides, with arginine being the preferred residue at position R2 (Figure 1.8A) [176]. Peptide-binding studies revealed that proper folding of HLA-B27 requires 30-fold higher number of peptides compared to HLA-A2, due to differences in B pocket residues [137]. Moreover, Dangoria *et al's* study has revealed that the disulphide intramolecular formation of newly synthesized HLA-B27 was dependent on the unpaired C67 [177]. Subsequent studies showed that unpaired C67 can contribute to the dimerization of HLA-B27 heavy chain, which has been proposed to be involved in AS pathogenesis [178, 179]. Lastly, B27.C67S transgenic rats develop disease with a milder phenotype than wild-type B27 transgenic rats [180]. However, so far, no study has confirmed that C67 is solely responsible for

dimerization or multimerization of HLA-B*27 heavy chains, due to the lack of other AS models containing variants at p67 of this molecule.

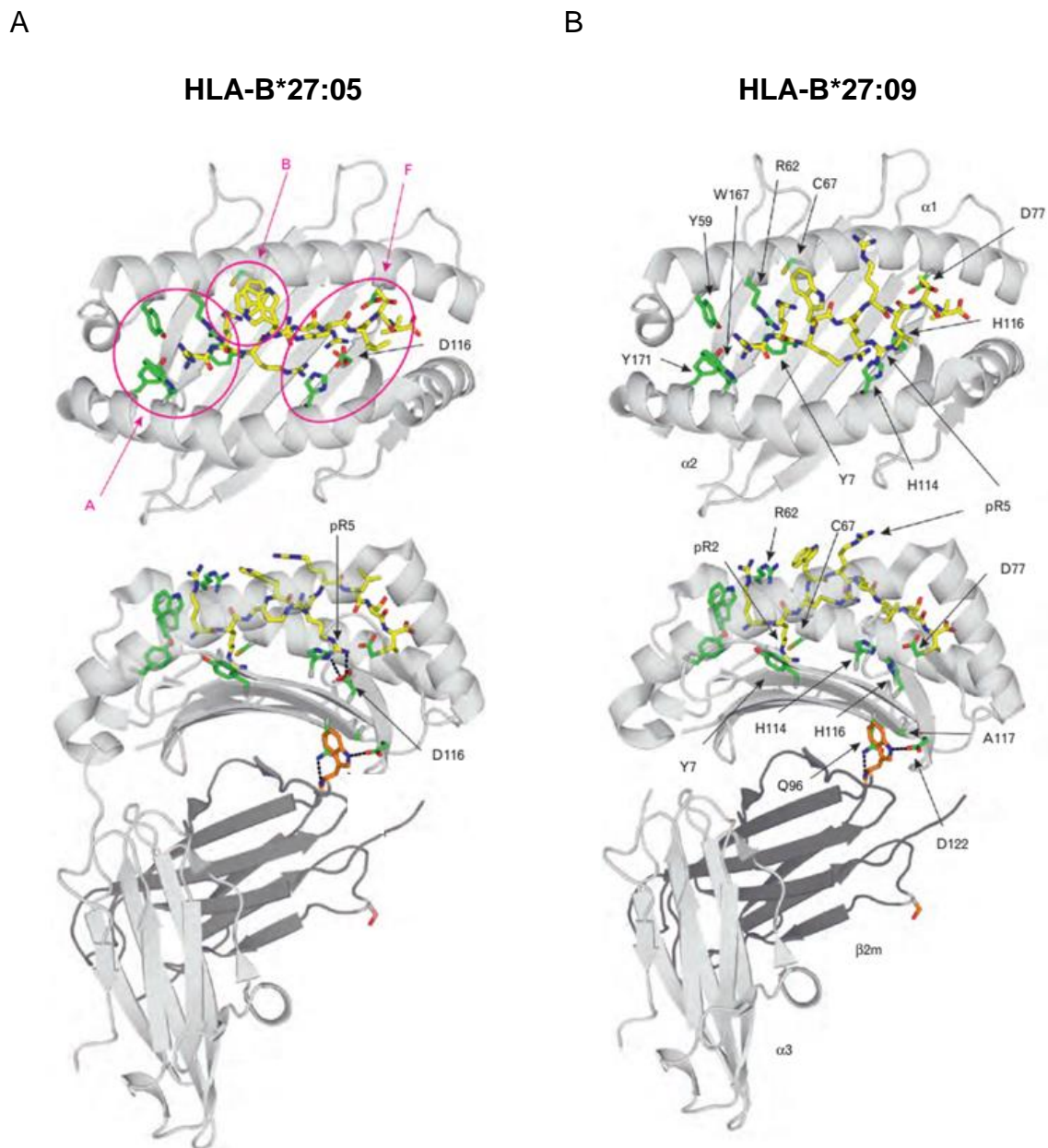


Figure 1.7: Structural features of HLA-B*27:05 and HLA-B*27:09.

Protein crystal structure of HLA-B*27:05 (A) and HLA-B*27:09 (B) are shown as viewed from the top and from the side of the α 2-helix. The heavy and light chains of each MHC complex are depicted in light grey and dark grey, respectively. For each MHC complex,

three binding groove pockets A, B and F are depicted, with each bound to an EBV-derived peptide pLMP2 ((RRRWRRLTV)(yellow). Nitrogen and Oxygen atoms are marked in blue and red, respectively. This illustration shows the effect of H116 on the conformation of bound pLMP2 peptide in HLA-B*27:09 (canonical-binding mode) versus D116 in HLA-B*27:05 (noncanonical-binding mode). The figure was adapted from Uchanska-Ziegler *et al* [123].

There is a single amino acid difference between HLA-B*27:05 (AS-associated subtype) and HLA-B*27:09 (non-AS-associated subtype) at p116 i.e. an aspartic acid (D) to histidine (H) change (Table 1.2). D116 is located on the floor of the binding groove which belongs to the F pocket which binds the peptide C-terminus. In HLA-B*27:09, H116 binds to the central pR5 of the peptide, pointing toward the solvent. In contrast, in HLA-B*27:05, D116 forms a salt bridge with the same peptide and binds to pR5 of the peptide but pointing into the binding groove [181]. As a result, although HLA-B*27:05 and B*27:09 may present a peptide with the same sequence, the distinct spatial structures can lead to different recognition by TCRs (Figure 1.8). Therefore, the amino acid at 116 could play an important role in presenting arthritogenic peptides and/or those that can potentially contribute to molecular mimicry. Further studies of D116 and H116 of HLA-B27 heavy chain may offer insights into the mechanism of AS and other related diseases.

1.5 Gene-editing Technology

Gene editing, the technology to introduce specific changes in the genome sequence of a living tissue or organism, has gone through many extraordinary milestones in its history of development. After the discovery of the DNA double helix structure in 1953 by James Watson, Francis Crick and Rosalind Franklin, scientists have continually been focusing on pursuing an effective method for DNA intervention and editing to treat human diseases. In 1968, Werner Arber discovered types of proteins which assisted bacteria to recognize and degrade bacteriophage DNA to protect the host which later became known as restriction enzymes [200]. The first gene-editing studies were conducted in yeast and mouse models during the 1970s and 1980s [201-204]. In 1985 and 1993, zinc finger nucleases (ZFN) and Clustered Regularly Interspaced Palindromic Repeats (CRISPR) were, respectively, discovered but their medical applications were still very limited at the time [205, 206]. Following the first human chromosome sequence in 1999, the 2000s observed the full Human Genome being mapped, which fueled the development of gene-engineering technology. One year after the discovery of the transcription activator-like effector nucleases (TALEN), in 2012 Jennifer Doudna, Emmanuelle Charpentier, and their research groups successfully revealed the biochemical mechanism of CRISPR activity, especially CRISPR-Cas9 [207]. One year later, the first use of CRISPR outside the world of bacteria was demonstrated by Feng Zhang and colleagues' whose protocol demonstrated manipulating genetic material in eukaryotic cells [208]. Zhang and colleagues successfully designed a CRISPR/Cas9 system to target the human Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1A (*DYRK1A*) at efficiencies of 65-68% for each locus in HEK cells. Additionally, targeting the Empty Spiracles

Homeobox 1 (*EMX1*) loci revealed that 49% of clones were heterozygous for the deletion, whilst 11% expressed a homozygous deletion [208]. In 2018, the CTX001 project was the very first CRISPR clinical trials of experimental treatment for genetic blood disorders which included severe sickle cell disease (SCD) and transfusion-dependent beta thalassemia (TDT). The project represented a collaboration between two companies, Vertex Pharmaceuticals and CRISPR Therapeutics in recruiting patients carrying mutations in the hemoglobin β subunit gene (*HBB*). The CRISPR-Cas9 system was designed to target the erythroid-specific enhancer region of BCL11A in hematopoietic stem and progenitor cells (HSPCs) isolated from patients by apheresis after mobilization. Disruption of BCL11A expression restored γ -globin synthesis and reactivated production of fetal hemoglobin. The latest data from the CTX001 project targeting TDT, showed that ten patients who had received gene-editing based therapy had made positive improvement in fetal hemoglobin levels, with no blood transfusions and pain relief required [209]. In the same year, Jennifer Doudna and Emmanuelle Charpentier were awarded The Nobel Prize in Chemistry for the development of CRISPR.

1.5.1 Mechanism of gene editing

Double-strand breaks (DSBs) in DNA are estimated to occur ten times per day per cell in dividing mammalian cells, which can lead to unwanted and deleterious chromosome rearrangements resulting in gene organization disruption and function [210, 211]. The outcomes of DSBs pose a potential threat to the stability of the genome, tumorigenesis, or risk premature cell death [212-214]. To minimize the damage caused by DSBs, different DSB repair pathways have evolved and are utilized by mammalian cells [212-214]. Each cell decides their DSB repair pathway based on their

cellular context which includes the local chromatin environment and cell-cycle phase [210]. In mammalian cells, the two-ended DSB can be repaired by following two main mechanisms: classical non-homologous end joining (NHEJ) and homologous recombination (HR) [215-217]. DSB at the target genomic sequence of interest is the first requirement for achieving high efficiency genome editing and NHEJ or HR are more favorable depended on the purpose of the application.

“Classical” NHEJ (cNHEJ) is distinct from the minor alternative end joining (aEJ) pathway which cells utilize to repair DSBs in the absence of cNHEJ proteins. CNHEJ is initiated by the binding of the dimeric protein complex Ku70/Ku80 to the broken DSB ends [218, 219]. Subsequently, the complex of DSB and Ku70/80 recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) to the broken ends and establishes a long-range synapse [220]. DNA-PKcs become activated via autophosphorylation and other proteins including the nuclease enzyme Artemis, which can trim DSB ends, are recruited to the complex of DSB, Ku70/80 and DNA-PKcs [221]. The heterodimer Ku70/Ku80 also recruits other NHEJ proteins including DNA ligase IV (LIG4), the scaffolding proteins X-ray cross-complementing protein 4 (XRCC4), XRCC4-like factor (XLF) and the paralog of XRCC4 and XLF (PAXX) [222-226]. XLF helps support the two DNA ends to become closely aligned, which enables LIG4, which is stabilized by XRCC4, to start ligating the long-range synapse. At the end of the NHEJ process, Artemis, DNA polymerases λ and μ and other enzymes bind to the short-range synapse and ensure both appropriate ligation and complementarity [227]. The final product of DSB repair following the NHEJ pathway is a DNA sequence with the ends being joined precisely with either local small insertions or deletions of nucleotides (indels). By introducing indel mutations, scientists often

utilize the NHEJ pathway to in-activate a specific gene of interest for research or clinical purposes.

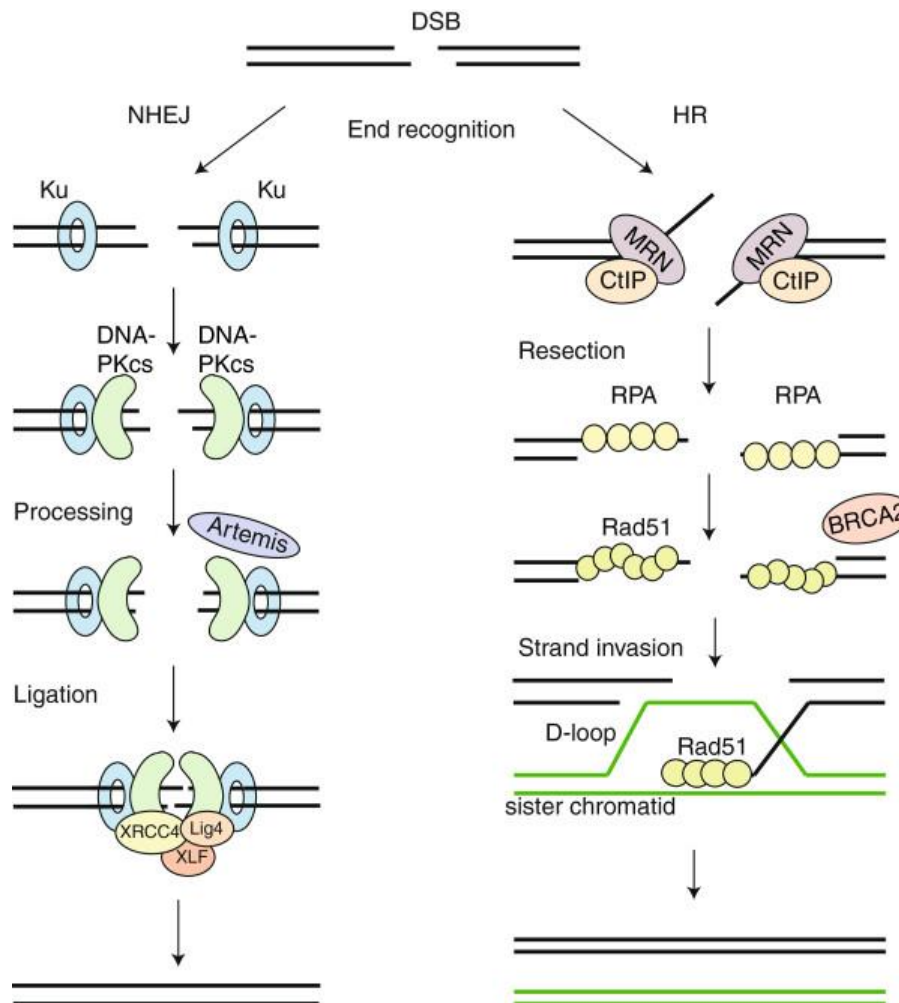


Figure 1.8. DNA double-strand break repair via non-homologous end joining and homologous recombination pathways.

The recognition of ends by the Ku70/80 complex initiates NHEJ, followed by the recruitment of DNA-PKs. Nucleases such as Artemis trim the incompatible DNA ends to 4-5 bases and then XRCC4-DNA Ligase IV-XLF ligation complex seals the break. The HR pathway involves re-sectioning on the DSBs, which is initiated by MRN-CtIP-complex to generate ssDNA which is later coated by RPA. Subsequently, BRCA2 promotes the replacement of RPA by Rad51 which forms a nucleoprotein filament on ssDNA and facilitates homology search and strand invasion, which lead to HDR. Illustration was obtained from Brandsma *et al* [228].

The second major pathway of DSB repair is Homologous Recombination (HR), in which a similar or identical sequence is used as a template for accurate repair. While cNHEJ operates throughout the vertebrate cell cycle, HR is restricted to S and G2 phases of the cell cycle only in higher eukaryotes [229]. The phase-restricted DR pathway was suggested to prevent the recombination with non-identical sequences and ensure error-free DNA replication of the sister chromatids. However, HR-mediated DSB repair can lead to loss of heterozygosity (if the correction takes place between maternal and paternal chromosomes) or the introduction of indel mutations (if repeats are not aligned properly). During the HR pathway, the re-sectioning of the broken DNA ends is initiated by the MRE11-RAD50-NBS1 (MRN) complex, which acts as a scaffold for activation of the Ataxia Telangiectasia Mutated protein (ATM), which is the main upstream kinase of DSB signaling [230]. Under ATM signaling, MRE11 endonucleases interact with C-terminal binding protein (CtBP) and nicks the strand that ends at a DSB to generate 3'-single stranded DNA (ssDNA) [231]. The secondary structure of the ssDNA is then removed by the activity of Replication Protein A (RPA) [232]. BRCA2, a breast and ovarian cancer predisposition gene, is subsequently activated by the replacement of RPA by RAD51 to form a nucleoprotein filament. The RAD51-ssDNA nucleoprotein filament promotes the recruitment of homologous sequences as a repair template. Based on the intact sequence, the 3'-ssDNA end is extended by the catalytic activity of RAD51 and other proteins including BRCA1, BARD1 and DNA polymerase δ (Pol δ) [233, 234]. After complete restoration of one strand, the 5' end of the broken DNA is captured and repaired via formation of two Holliday junctions, where each junction represents a branched nucleic acid structure containing 4 double-stranded arms, to form the identical sequence with the template [228].

1.5.2 Classification of Gene-editing Technology

The development of genome editing technology has been accompanied by the discovery of programmable nuclease systems which have been used to introduce site-specific modifications into genomic DNA by producing DSBs. There are four major nuclease systems which include Meganucleases (MegNs), Zinc-finger nucleases (ZFNs), Transcription Activator-like Effector Nucleases (TALENs) and more recently Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) (CRISPR/Cas-9) (Figure 1.10). Although these nucleases share the ability to generate DSBs at specific DNA regions of the host chromosomes, each of the systems still have unique characteristics.

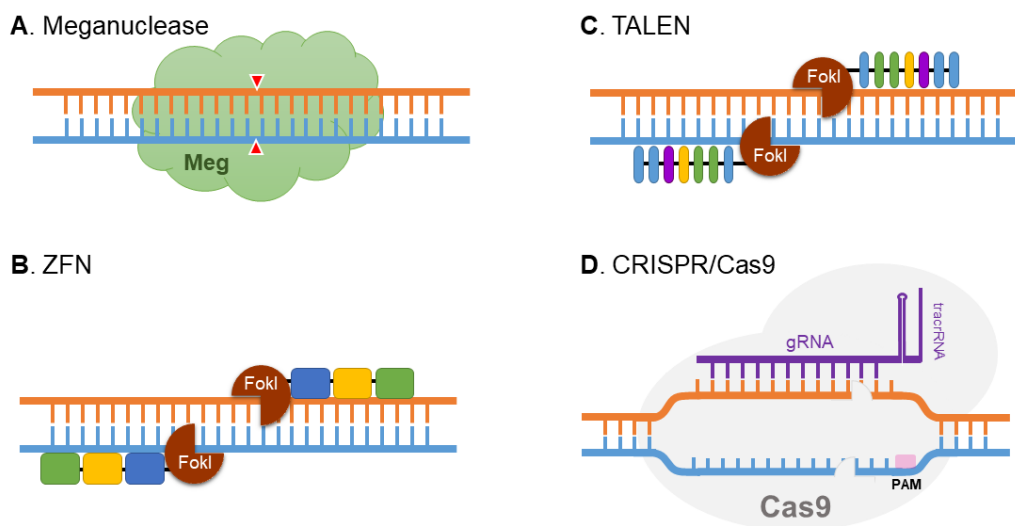


Figure 1.9. Schematic diagram of four gene-editing endonuclease systems.

(A) Meganucleases (MegNs) contain long DNA-binding recognition sites (14-40 nucleotides). (B) Zinc-finger nucleases (ZFNs) consist of the restriction endonuclease (Fok1) and zinc-finger domains recognizing 3 base pairs. (C) Transcription Activator-like Effector Nucleases (TALENs) are composed of a DNA-cleaving enzyme (Fok1) and array of nuclear effector domains which recognize single base pairs. (D) Clustered Regularly

Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) (CRISPR/Cas-9) is a complex of Cas9 nuclease and an artificially designed single-guide RNA (sgRNA). The sgRNA navigates the complex to the target DNA site and Cas9 nuclease generates DSBs. The illustration were obtained from Alstem [235].

Families of Meganucleases (MegNs) are often encoded by genes located within self-splicing elements and found in eukaryotic mitochondria and chloroplasts. MegNs have an important role in exhibiting lateral mobility of genetic elements within an organism and are known as “homing” endonucleases, which can cleave DNA at specific sites and facilitate DSBs reparation using the HDR pathway followed by integrating MegN-encoding sequences at new sites [236]. MegN consists of a DNA-binding domain comprising 14-40 bp and a DNA-cleavage domain which can cleave dsDNA. Long recognition sites by MegNs bring a huge advantage to therapeutic applications by significantly increasing the specificity of DNA targeting compared to other genetic editing technologies [237]. However, some biological characteristics of MegNs result in many drawbacks in their use for clinical applications. First, MegNs only introduce DSBs at known cleavage sites so operators must initially introduce this site into the region of interest. Secondly, two functional domains of MegNs, including DNA-cleavage domains and DNA-binding domains, are not separated distinctly so it is difficult to design the sequence-specific enzymes for all desired targets. Thus, MegNs present a challenge in gene-editing applications due to its complexity, expense, and the time-consuming process.

With respect to transcription factors, Zinc finger nucleases (ZFNs) have been artificially engineered to become site-specific restriction enzymes for the purpose of genome editing. ZFNs are chimeric molecules, where each subunit contains two

domains: the Fok1 endonuclease and zinc finger domains (Figure 1.10B). The restriction endonuclease Fok1, which originated from *Flavobacterium okeanoicoites* [238], consists of a DNA-binding domain and a non-specific DNA cleavage domain at N- and C- termini, respectively [239]. The Fok1 nuclease domain is fused to the zinc finger domains, which are transcription factors, and each finger recognizes specific sequence of 3-4 bases. Several studies have shown that higher numbers (3-6) of zinc fingers enhance the efficiency and specificity of ZFNs introducing DSBs [240, 241]. Based on the known patterns recognized by preexisting zinc fingers, the ZFNs structure is easy to design by using computational assembly. However, the cleavage specificity of ZFNs is “context-dependent” because it does not only depend on the target sequence but also on adjacent sequences [236]. This limitation can lead to genome instability and fragmentation in the case where many non-specific cleavages occur. Additionally, ZFN technology only facilitates targeting one single target at any one time and can cause toxicity to host cells due to off-target effects [242]. The insufficient specificity of DNA binding domains or inappropriate interactions between domains facilitate DSBs at similar target sequences within the genome.

Similarly to MegNs and ZFNs, Transcription Activator-like Effector Nucleases (TALENs) are also artificially engineered restriction enzymes but they tend to be more efficient, specific and less expensive [243]. TALENs are similar to ZFNs in that they are also heterodimeric proteins consisting of a cleavage and DNA-binding domain. While they both use the Fok1 restriction endonuclease as the cleavage domain, only the DNA-binding domain of TALEN is composed of highly conserved repeats derived from transcription activator-like effector, which is from the *Xanthomonas* plant pathogenic bacteria [244]. DNA-binding domains of TALENs are capped by a transport

and nuclear localization signal at the N- and C- termini, respectively. The middle region of TALEN is a repeating 33-34 amino acid sequence with a variable region at the 12th and 13th amino acid, termed Repeat Variable Di-residue (RVD). A combination of appropriate RVD regions allows specific TALEN binding to its target sequence [245]. Compared to ZFNs, TALENS are easier to design, exhibit greater specificity, more efficient in producing DSBs and are less toxic to host cells [246]. However, TALENS technology is limited due to their large size, with the size of approximately 3kb for cDNA coding a TALEN cassette, leading to difficulty with the delivery of the molecules into cells for clinical application [247].

The RNA-guided engineering nucleases, known as the CRISPR/Cas system, is the most recent platform developed for performing gene editing. Although the clustered regularly interspaced short palindromic repeat, also known as “short regularly spaced repeats”, was first discovered in the 1980s, this technology had to wait more than three decades to be exploited intensively as a promising tool for precisely editing genomes. Computational analysis showed that short regularly spaced repeats were found in more than 40% of bacteria and 90% of archaea [248, 249]. These elements were later known as CRISPRs. A group of genes were also identified to be located adjacently to the CRISPR locus, which encoded the “CRISPR-associated system” or Cas proteins [248]. The CRISPR/Cas system was suggested to play vital roles in the immune system of bacteria and archaea against invading phage and plasmids [250]. While ZFNs and TALENS recognize their target via DNA-protein binding, CRISPR/Cas systems are guided by RNA-DNA interactions. All CRISPR/Cas systems consist of two components: a cas protein and a guide RNA (gRNA or sgRNA). There are different variants of Cas proteins such as Cas6 [251], Cas9, Cas12a or Cas12b [252] but my

study was only focused on the type II Cas9 nucleases which are also the most intensively used in biological and biomedical research. Compared to other gene-editing technologies, the advantages of CRISPR/Cas systems are their relative simplicity, great efficiency, and flexibility for different applications. Details of CRISPR/Cas9 components are discussed in the next section.

1.5.2 CRISPR-based technologies

1.5.2.1 Natural CRISPR/Cas9 systems

CRISPR/Cas9 systems are the best characterized platforms used for mammalian genome editing. There are several CRISPR/Cas9 platforms originating from different species of bacteria and archaea, which are all composed of Cas9 nucleases and sgRNA. The sgRNA is a chimeric molecule composed of the CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) (Figure 1.6A). The CrRNA array is encoded by short variable sequences derived from exogenous DNA targets, which are known as protospacers [253]. Each crRNA contains 20 nucleotides to guide Cas9 to the DNA target via Watson-Crick base pairing and a partial direct repeat which is complementary with part of the tracrRNA. Beside pairing with the crRNA, tracrRNA has an additional long stretch of bases forming a “stem loop” structure which is bound by the Cas9 nuclease. Recently, the sgRNA duplex has been simplified to one single RNA molecule by fusing crRNA and tracrRNA together (Figure 1.11A) [254].

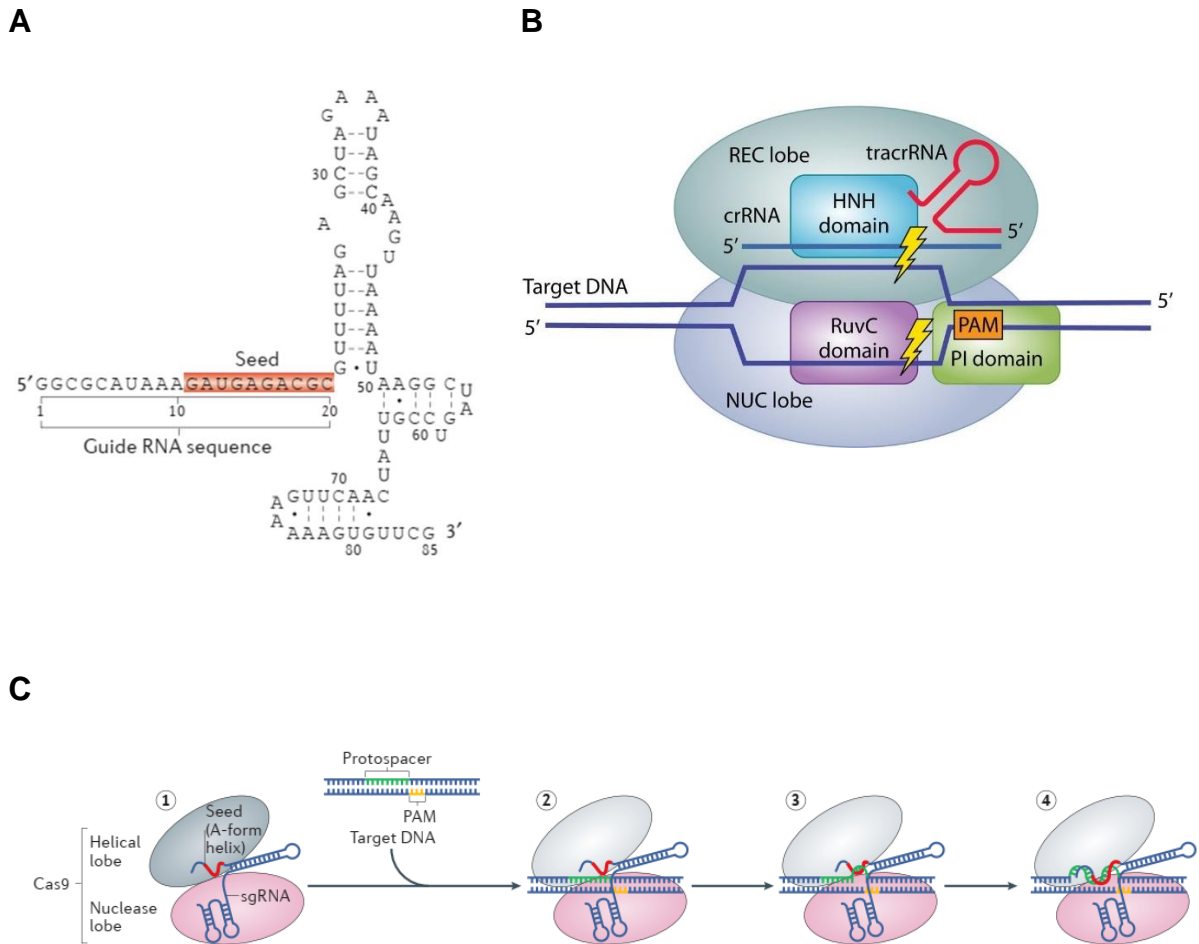


Figure 1.10. Schematic highlighting components of sgRNA and Cas9 nuclease.

(A) Detailed structure of the sgRNA which contains the 20-nt guide nucleotide and the stem loop RNA scaffold. (B) Cas9 protein consists of two lobes: the recognition (REC) and nuclease (NUC) lobes. The REC lobe is responsible for the formation of the RNP with the sgRNA duplex. The NUC lobe consists of two nuclease domains (HNH and RuvC) and a C-terminal region containing a Protospacer Adjacent Motif (PAM) - interacting (PI) domain. (C) Schematic of Cas9-crRNA complex targeting DNA. (1) After forming the complex of CRISPR/Cas9, the crRNA seed sequence (red) is available for pairing in an 19-20 nt A-form helix; (2) CRISPR/Cas9 complex binds to target DNA after recognizing the PAM sequence (yellow) on the non-complementary strand; (3) Interaction between Cas9 and the DNA duplex initiates strand separation on the protospacer sequence of target DNA (green). (4) Cas9 nuclease performs cleavages under the propagation of helix formation. Figure is adapted from Gorski *et al* [255, 256].

The CRISPR/Cas system uses Protospacer Adjacent Motifs (PAMs), which are DNA motifs located within the target site. Each type of Cas9 nuclease can be distinguished by specific properties related to size, location of the introduced DSB and PAM sequence requirement (Figure 1.12). Among the different types of CRISPR/Cas9 systems, *Streptococcus pyogenes* Cas9 (SpCas9) protein is commonly used, and it is made up of 1368 amino acids with a molecular weight (MW) of ~160 kDa. The PAM motif of SpCas9 is 5'-NGG-3', in which "N" can be any nucleotide. When the sgRNA and PAM motif are matched with the target sequence, spCas9 introduces the DSB between nucleotides at position 3 and 4 upstream of the PAM sequence. The Cas9 nuclease consists of two nuclease domains: HNH and RuvC. The wild type spCas9 introduces DSBs by using both nuclease domains. "Nickase" Cas9 is modified to inactivate HNH domains by modifying histidine to alanine (H84A) or the RuvC domain by modifying aspartate to alanine (D10A) to create single stranded breaks only, which are also known as "Cas9 Nickase" [257-259]. CRISPR-Cas9 nickase-based systems, which require 2 separate sgRNAs targeting opposite DNA strands in close proximity, can be employed to generate a staggered DSB with overhangs [260]. Therefore, using Cas9 nickase can achieve higher knock-in efficiency and lower off-target effects.

Enzyme name	Size (residues)	PAM requirement and cleavage pattern
SpCas9/FnCas9	1368/1629	
St1Cas9	1121	
St3Cas9	1409	
NmCas9	1082	
SaCas9	1053	
AsCpf1/LbCpf1	1307/1228	

Figure 1.11. Property Comparison of Natural CRISPR/Cas9 Systems.

Each strain of bacteria has a different type of Cas9 variant which are distinguished by size, PAM motifs and length of gRNA-binding regions. Variants of CRISPR nuclease are different in size fluctuating between 1053 and 1629 amino acids and PAM motif requirement (blue sequence). The Cas9-mediated cleavage sites are pointed by the red arrows. Table was adapted from Komor *et al* [261].

1.5.2.2 Delivery system of CRISPR/Cas9 for genome editing

Since the discovery of the CRISPR/Cas9 system, this powerful gene-editing technology has been applied to various approaches including disease modelling transcriptional regulation, nucleic acid imaging and gene correction. Especially, CRISPR/Cas9 has intensively revolutionized the field of gene therapy by improving the merits of accuracy and simplicity. However, one of the foremost barriers of this technology is the insufficiency of rational delivery. Therefore, different CRISPR/Cas9

platforms has been developed to overcome this difficulty. This section highlights the design of CRISPR/Cas9 systems in forms of viral and non-viral particles including plasmids, rRNAs and ribonucleoproteins (RNPs).

Up to the present day, delivery methods for clinical trials of *in vivo* gene-editing by the CRISPR/Cas9 system have been mostly based around the transduction of viral vectors including adenovirus (AV), lentivirus (LV) and adeno-associated virus (AAV) [262]. All viral vectors have been modified to prevent self-replication or being inherited by daughter cells. Each type of viral vector has its own advantages and disadvantages which are summarized in table 1.4. AV is a non-enveloped virus with an icosahedral nucleocapsid encapsulating the viral dsDNA genome. The high packing capacity (>8 kb) allows AV to deliver Cas9 and multiple sgRNA sequences for multiple targets for gene-editing. For instance, AV-mediated CRISPR/Cas9 was employed to excise a 23kb genomic region covering mutant exon 23 of Duchenne Muscular Dystrophy (*DMD*) to block the disruption of dystrophin in muscle fibers in a mouse model of DMD [263]. In another study, an AV containing Cas9 and sgRNA targeting the missense mutation (CTG > CGG) in the Epidermal Growth Factor Receptor (*EGFR*) oncogene were employed to disrupt the mutant allele and enhance cancer cell killing in a xenograft mouse model of human lung cancer [264].

Table 1.4 Characteristics of viral vectors in clinical application.

Type	Size limitation	Advantage	Disadvantage	References
AV	>8 Kb	No-integration, large capacity	High immunogenicity	[271]
AAV	4.5 Kb	Stable transgene expression, serotype-specific targeting, low immunogenicity	Low capacity	[272]
LV	8 Kb	High capacity, mild immunogenicity, stable expression	Long lasting expression of Cas9	[273]

AAVs are small non-enveloped viruses containing linear single-stranded DNA (ssDNA) genomes of approximately 4.8 Kb. Due to the low immunogenicity, serotype-specific targeting, and stable transgene expression, AAVs have become the most preferable choice for gene therapy. However, the packing size of AAVs is limited at around 4.5 Kb, so small cas9 nucleases including truncated SpCas9 or *S.aureus* Cas9 (SaCas9) have been used to solve this issue [265]. For example, CRISPR/Cas9 mediated AAV has been used to introduce gene corrections of adult phenylalanine hydroxylase (*Pah*), which is associated with the human autosomal recessive liver disease phenylketonuria (PKU) [266]. The outcome of this study reported that up to 63% of the mRNA exhibited a corrected sequence with PAH enzyme activity restored in a mouse model for PKU.

Another potential delivery system for CRISPR/Cas is employing HIV-1 derived LV, a type of simple retrovirus. LV is a single-stranded RNA spherical virus using its glycoprotein envelop to attach and enter the host cell. Owing to many prominent merits

including high capacity (8 Kb), mild immunogenicity and stable expression, LV is an excellent delivery method of CRISPR/Cas9 into nondividing cells, which have no ability to pass the integrated vector to daughter cells. Due to the ability of host genome integration, off-target effects caused by LV-mediated delivery are greater than other approaches such as AV or AAV. To overcome this issue, a self-inactivating Cas9 system was developed by using two sgRNAs: one for the target gene and one for disrupting the Cas9-coding gene [267]. Additionally, Angelo *et al* [268] constructed the integration-defective LV (IDLV) by introducing point mutations in the integrase protein of LV. IDLV-mediated transduction of CRISPR/Cas9 was exploited successfully *in vitro* and *in vivo* to modulate expression levels of the *SNCA* gene, which encodes a small protein called alpha-synuclein, in dopaminergic neuron-derived induced pluripotent stem cells (hiPSC) [269]. In summary, viral vector transduction of CRISPR/Cas9 is the preferable choice for gene-editing therapy due to their superior transfection efficiency and general application [270]. On the other hand, viral delivery systems require improvements with respect to specificity to lower off-target effects, immunogenicity and increase the packaging capacity.

While viral approaches are more suitable for *in vivo* genome editing, non-viral nanovectors for CRISPR/Cas9 delivery present a promising strategy for basic research and *ex vivo* gene therapy. Compared to viral platforms, non-viral nanovectors have some advantages including causing less cell injury, higher packaging capacity and reduced risks of immune activation. Three major non-viral forms of CRISPR/Cas9 delivery are plasmid Cas9 DNA vector (pDNA), Cas9 mRNA and Cas9 ribonucleoprotein (RNP) (Figure 1.13).

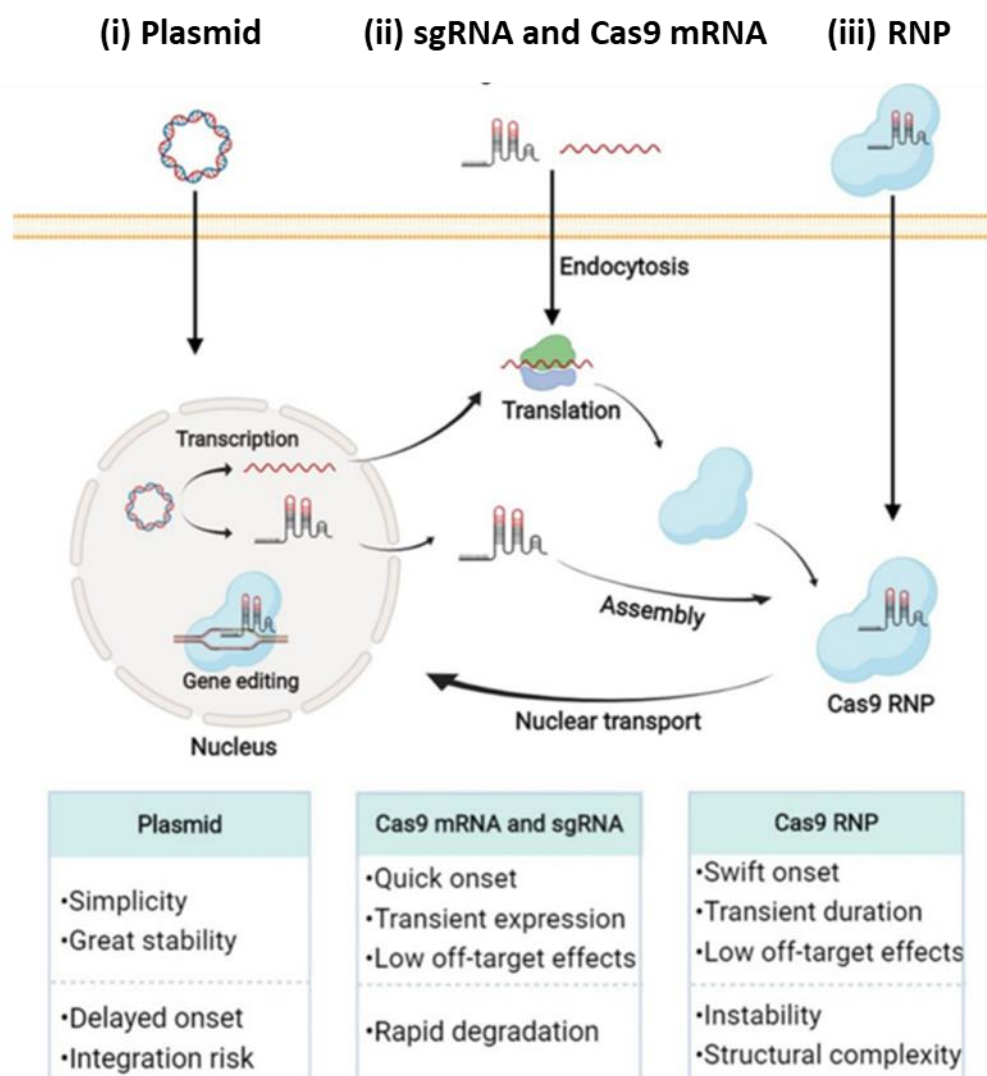


Figure 1.12. Schematic diagram of features of CRISPR/Cas9 platforms including plasmids, mRNAs and RNPs.

Before being transported into the nucleus, Nuclear Localization Signal (NLS)-fused Cas9 RNP can appear in the cytosol in three different ways: (i) transfection of Cas9 and sgRNA DNA plasmid followed by the transcription and translation of the platform; (ii) transfection of sgRNA and Cas9 mRNA which are translated into Cas9 RNP in the cytosol; (iii) direct transfection of RNP containing Cas9 protein and sgRNA. The table below shows advantages and disadvantages of each platform. Illustration was obtained from Cheng *et al* [270].

Using CRISPR/Cas9 encoding plasmids which were developed by Feng Zhang's group, the delivery of gene-editing tools can be rapid [274]. Two expression cassettes (pX330) can be used containing the Cas9 protein gene sequence and gRNA separately. Within the Cas9 cassette, SpCas9 protein was expressed under the control of a chicken β -actin promoter and fused with the nuclear localization signal (NLS) for targeting the CRISPR/Cas9 system into the nucleus. The gRNA cassettes can be of two forms: accessible crRNA with a core tracrRNA, or a complete single guide RNA (sgRNA). Both types of gRNA vectors require the U6 promoter, a type III RNA polymerase III promoter, to drive the expression of RNA. Although a single expression cassette containing all CRISPR/Cas9 components including Cas9, gRNA and HDR template were designed, this large plasmid was difficult to encapsulate [275]. CRISPR/Cas9 pDNA vectors are delivered into cells by lipofection using polyethylene glycol (PEG), polyethyleneimine (PEI) or exosomes [275, 276]. For instance, CRISPR/Cas9 plasmids, targeting the tumor suppressor gene MTH1, were delivered by PEI-derived fluorinated polymer (PF33) and gave a high transfection efficiency as well as enhanced MTH1 downregulation in mice [277]. The CRISPR/Cas9 pDNA is a promising candidate for gene-editing research, due to its stability, simplicity, and low cost of manipulation. However, limitations of CRISPR/Cas9 plasmids include difficulties in compacting the cassette with large sizes (>10kb) into nanoparticles and translocation of the systems into nuclei [278].

To overcome the disadvantages associated with pDNA CRISPR/Cas delivery, an RNA-based CRISPR system has been tested and is now a potential approach for *in vivo* genome editing. Basically, after entering the cytoplasm, Cas9 mRNAs are translated into Cas9 protein, and then assembled into RNP with transfected sgRNAs

followed by nuclear transport for genome editing. The advantages of RNA-based CRISPR systems is that they are smaller molecular structures compared to pDNA, have a quick onset, in which the transcriptions of Cas9 and sgRNA are skipped, and low level off-target effects. The poor stability of RNA is one of the main limitations with this CRISPR/Cas9 system and therefore requires methodology to protect the RNA during delivery. A lipid nanoparticle (LNP)-based carrier, known as LAP-INT01, have been exploited to compress the structure and stabilize Cas9 mRNA and sgRNA into a single nanoparticle [279]. Based on LAP-INT01, dendrimer-based lipid nanoparticles (dLNPs) have been developed to encapsulate multiple components for HDR including Cas9 mRNA, sgRNA and donor templates in a ssDNA form. In Farbiak's study [280], dLNP were employed to deliver Cas9 mRNA, sgRNA, and donor template into BFP/GFP switchable HEK293 cells with a single amino acid mutation Y66H. Fluorescence imaging, DNA sequencing and flow cytometry results showed that HDR efficiency was 55% *in vitro* and approximately 20% in xenografted tumors *in vivo*. Most recently, extracellular vesicles (EVs), which are lipid bound vesicles secreted by cells into the extracellular space, have also been applied to package and deliver RNA-based CRISPR/Cas9 systems due to their high biostability, high efficiency in uptake by cells and low cytotoxicity [281].

The last non-viral vector for CRISPR/Cas9 delivery is the use of the Cas9 RNP platform in which the expression of Cas9 protein and sgRNA could be skipped. In this method, purified recombinant Cas9 protein is fused with a NLS peptide and combined with chemically modified sgRNA to form Cas9 RNP complexes. These Cas9 RNP sgRNA complexes are then delivered directly to cells for introducing DSBs within desired sequences. By skipping the need for intracellular transcription and translation,

Cas9 RNP delivery can minimize the risk of off-target effects and genome integration [282]. Besides, Cas9 RNP complexes do not have long half-lives, estimated to be ~10.6 hours in HEK293T cells, so it is possible that this platform can avoid initiating immune responses and is suitable for genome-editing within stem cell lines including tissue stem cells, induced pluripotent cells, and embryonic cells [283]. However, the disadvantage of using Cas9 RNP complexes are their large size, low efficiency of endosomal escape and inherent instability, which requires optimal delivery methods (Figure 1.14)

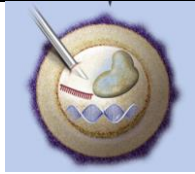
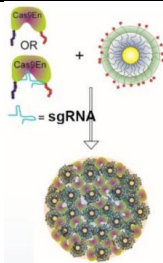
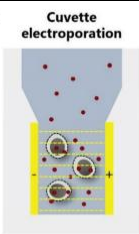
Cas9 RNP delivery platforms	Microinjection [284]	Lipid nanoparticle [285]	Electroporation [286]
			
Advantages	<ul style="list-style-type: none"> • No limit on molecular weight 	<ul style="list-style-type: none"> • High efficiency in gene-knock-out 	<ul style="list-style-type: none"> • Eligible to deliver donor template into nucleus • High efficiency in gene-knock-in
Disadvantages	<ul style="list-style-type: none"> • Require high skilled manipulation • Expensive 	<ul style="list-style-type: none"> • Difficult to deliver donor template into nucleus • Low efficiency on suspension cell 	<ul style="list-style-type: none"> • Low cell viability on sensitive cell

Figure 1.13. Cas9 RNP delivery routes and their features.

Methods for RNP delivery into mammalian cells include microinjection, lipid nanoparticles and electroporation. Microinjection is a physical method which directly injects foreign materials into living cells using micropipettes. Microinjection of Cas9 RNP has been successfully implemented in various cell lines, especially pluripotent stem cells like human embryos. For instance, in Ma's study [284], a CRISPR-Cas9 system targeting *MYBPC* mutations, which cause hypertrophic cardiomyopathy (HCM), were injected into human zygotes produced by fertilizing oocytes from a healthy donor (*MYBPC3^{WT/WT}*) with sperm from a patient heterozygous for the *MYBPC3* mutation (*MYBPC3^{WT/ΔGAGT}*). Sequencing of blastomeres collected from 54 CRISPR/Cas9-injected embryos revealed that only 9.3% were heterozygous *MYBPC3^{WT/ΔGAGT}*, while 66.7% were homozygous for the wild-type allele *MYBPC3^{WT/WT}* and 33.3% were uniform or mosaic heterozygous. This result indicated that injection of a CRISPR/Cas9 system was feasible for introducing gene modification in human zygotes [287].

Cationic lipid nanoparticles are the most common method for transfecting exogenous genetic material into cells [288]. Due to the negative charge of Cas9 nuclease and sgRNA, these molecules can be taken up and delivered into mammalian cells by commercial cationic lipid nanoparticles including Lipofectamine 2000 and Lipofectamine 3000 (Thermo Fisher). The lipofection of Cas9 RNP has been used to *in vivo* edit multiple pathogenesis-related genes for treating non-genetic degenerative disease [289]. However, lipid particles are only able to infiltrate material through the cell's plasma membrane, so they are not highly efficient in delivering DNA donor templates into nuclei for DSB repair following HDR. Therefore, Cas9 RNP complexes

must be engineered to bind with donor templates and then use the NLS to efficiently target the whole complex into the nucleus [290].

Electroporation is a delivery method using high voltage pulses to disturb the phospholipid bilayer of membranes where temporary nanopores are subsequently introduced which allows exogenous materials to be taken up by the cell. Electroporation can deliver both Cas9 RNP and donor templates to the nucleus and thus facilitating DSBs and HDR which are required for precise gene correction. The major drawback of CRISPR/Cas9-delivery via electroporation is substantial cell death caused by the high electrical pulse [291].

1.5.2.3 Donor template for gene correction

The most obvious difference of knock-in compared to knock-out is the requirement of exogenous repair templates or donor templates. Beside the desired insertion sequence, donor templates also contain the homologous arms which are complimentary with the upstream and/or downstream regions of the DSB site. Additionally, donor templates can be delivered into a cell, most often in one of three forms including Adeno-associated virus (AAV), double stranded DNA (dsDNA) and single stranded DNA (ssDNA). AAV-repair templates have been shown to provide extremely high knock-in efficiencies due to natural viral capabilities such as protecting DNA from degradation or stimulating homologous recombination [292]. However, compared to dsDNA and ssDNA, the drawbacks for AAVs are that they are relatively time consuming and expensive to produce and limited in the length of donor template due to the small packaging capacity (<5kb) of the virus. In contrast, dsDNA can be produced in the form of plasmid, which is feasible to make in any biological lab and can be much larger in size than AAVs. Nevertheless, plasmid DNA increases the off-

target effects by accidentally integrating into host genomes at CRISPR cut sites or even at random regions. Additionally, dsDNA mediated CRISPR edits have relatively low efficiency and higher toxicity compared to other types of repair template [293].

In mammalian cells, plasmid DNA or AAV donor template with at least 1-2 kb of total homology is frequently used to introduce large nucleotide changes but ssDNA sequences usually provide higher efficiency and cell viability for small changes [294]. These small changes include single nucleotide polymorphisms (SNPs), amino acid substitutions or epitope tags. The most common type of ssDNA used for HDR experiments is to use donor oligonucleotides (ssODN) which contains the desired change (~50nt) and homology arms varying from 30-90 nucleotides in length. DNA modifications such as Affinity Plus DNA Oligonucleotides (IDT) are usually applied during synthesis to increase the stability and exonuclease resistance of ssODN. Beside the desired changes, donor templates are recommended to include a restriction enzyme site to facilitate detection of successful HDR by restriction fragment length polymorphism (RFLP) analysis, which is a combination of PCR and restriction enzyme digestion analysis.

1.6 Application of Gene-editing Technology in Immunology

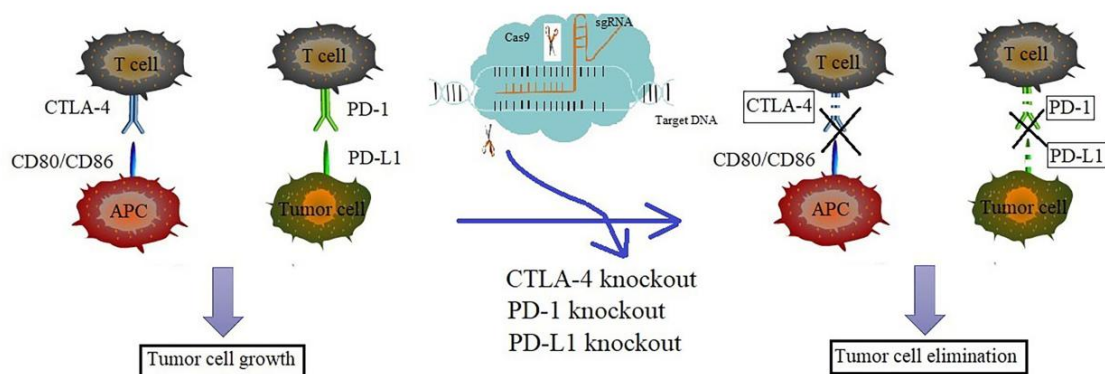
1.6.1 General application in immunology

The development of CRISPR/Cas9 has brought many useful applications in immunological research and the treatment of diseases related to the immune system. In the field of immunotherapy, the CRISPR/Cas9 system is not only considered as a powerful gene-editing tool for identifying novel biomarkers, drug targets and drug resistance genes but has also become a promising therapeutic approach in cancer immunotherapy [295].

CRISPR/Cas9 screening is a high throughput genetic loss-of-function experimental method for studying gene function or finding disease related genes. Scientists have used CRISPR/Cas9 libraries containing more than 120,000 sgRNAs to generate tumor cells with multiple genomic lesions followed by exposing to specific CTLs [296]. The distribution of the sgRNA reads were analyzed by deep sequencing to identify the sgRNA library representation in tumor cells before and after co-incubation with T cells. Beside MHC-I related genes such as *HLA*, *B2M*, *TAP1*, *TAP 2* and *TAPBP*, the results demonstrated that previously undescribed genes, such as the cellular signaling-related genes *APLNR* and *BBS1*, and microRNAs were contributing to promote T cell damage to tumors. Via CRISPR screening, researchers have also studied more deeply the mechanism of T cell activation and identified a novel candidate for tumor immune drug intervention, known as *CYRI-B* or *FAM49B* [297]. Deficiency of either the *CYRI-B* or *FAM49B* genes led to hyperactivation of Jurkat T cells, an immortalized line of human T lymphocyte cells, following T cell receptor stimulation as demonstrated by p21-activated kinase (PAK) phosphorylation, CD69 induction and actin assembly which is crucial for cell–cell adhesion, cell motility and epithelial differentiation [298].

CRISPR/Cas9 has also been applied in the development of therapies to overcome immune tolerance to weaken cancer immune resistance via blocking immune checkpoints including Programmed cell death protein 1 (PD-1), Programmed death-ligand 1 (PD-L1) and cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) (Figure 1.15A) [299]. PD1/PD-L1 and CTL-4 expressed on the surface of some immune cells, function to regulate T cell responses. When PD-1, an immune checkpoint found on the surface of T and B cells, is bound to its ligands including PD-L1 and PD-L2, it prevents T cells from killing other cells. Tumor cells can take advantage of these immune checkpoint proteins to evade or inhibit CTLs. The blockage of PD-1/PD-L1 or CTLA-4 pathways could enhance the efficiency of T cell mediated immunotherapy targeting cancers [300]. CRISPR/Cas9 technology has been used to knocked out PD-1 in CTLs, with the aim of enhancing the cytotoxic effect of CD8⁺ T cells against tumour targets [301]. The results showed that PD-1-KO CTLs secreted higher levels of TNF- α and IFN- γ and enhanced tumor cell death by 40% compared to wild type cells. The xenografted mice transfused with PD-1-KO CTLs also had longer survival compared to the control group [301]. Similarly, CRISPR/Cas9-mediated CTLA-4 disruption led to increased cytokine secretion and enhanced T cell responses against the CD80/CD86-positive bladder cancer cells [302]. The study by Yahata and colleagues used CRISPR/Cas9 to knock out PD-L1 in a murine ovarian cancer which were subsequently inoculated into syngeneic mice for evaluating survival and tumor dissemination [303]. The results demonstrated that CRISPR/Cas9-mediated PD-L1 disruption in ovarian cancer cells enhanced anti-tumor immunity by increasing tumor infiltrating lymphocytes and regulating cytokine/chemokine profiles in the tumor microenvironment.

A



B

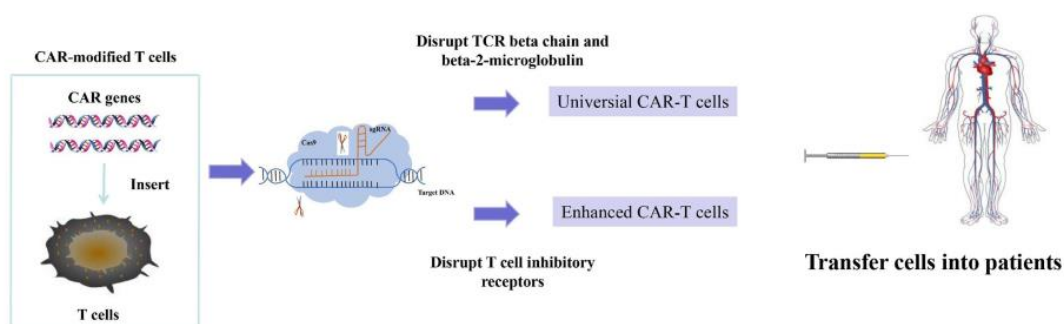


Figure 1.14. The application of CRISPR/Cas9 in immunotherapy.

(A) CRISPR/Cas9-mediated KO of PD-1/PD-L1 or CTLA-4 enhanced immune responses against tumor cells. (B) The combination of CRISPR/Cas9 technology and chimeric antigen receptor T cell (CAR-T) therapy. CRISPR/Cas9 has been exploited to disrupt the TCR beta chains, B2M and PD-1 to generate universal CAR-T cells and enhance their effect against tumor cells.

One of the most recent approaches of anti-tumor immunotherapy is the combination of CRISPR/Cas9 technology and chimeric antigen receptor T cell (CAR-T) therapy (Figure 1.15B). CAR is a protein construct consisting of an intracellular signaling domain, a transmembrane domain, a spacer domain, and a single-chain variable fragment (scFv) which is specific to tumor-associated antigens (TAA) on cancer cells [304]. CAR-T cell therapy is a promising treatment for B-cell malignancies, especially

in acute lymphoblastic leukemia (ALL) in which the complete remission rate of this therapy is as high as 90% [305]. One of the limitations of CAR-T therapy is the lack of qualified autologous T cells from cancer patients. To overcome this difficulty, scientists have focused on generating off-the-shelf or universal CAR-T cells from healthy donors. Therefore, CRISPR/Cas9 systems have been exploited to disrupt B2M and the TCR beta chain on donor CAR-T cells [306]. While disruption of B2M can prevent the expression of MHC-I on the surface of allogeneic T cells and limit the recognition by host immune cells, disruption of the TCR beta chain can inhibit the expression of endogenous TCR. Furthermore, CRISPR/Cas9-mediated disruption of PD-1 can block T cell inhibitory receptors, as described previously, to improve function of CAR-T cells against tumor cells. Application of triple-KO (TCR, B2M, and PD-1) allogeneic CAR-T cells via CRISPR/Cas9 has been reported to enhance anti-tumor *in vivo* activity without significant graft rejection [306, 307]. Moreover, transforming growth factor beta (TGF- β) receptor II (TGFB2) is another candidate for CRISPR/Cas9-mediated knock-down to increase efficacy of CAR-T therapy for the treatment of solid tumors [308]. Thus, CRISPR/Cas9 gene editing technology has made a massive contribution in optimizing CAR-T cell therapy to adapt to the complexity of cancer biology and continues to establish new approaches for cancer treatment.

1.6.2 Targeting the MHC-I Complex by CRISPR/Cas9 technology

Playing a central role in the regulation of the immune system, MHC complexes have become a preferred subject of interest for gene editing technology. Since the discovery of CRISPR/Cas9, this technology has been applied to several studies disrupting MHC-I expression (Table 1.5). CRISPR/Cas9-mediated disruption of the MHC-I complex has been facilitated by knocking out B2M or HLA class I heavy chain genes. The main purposes for MHC-I disruption were to generate cell models for studying mechanisms of peptide antigen processing or produce universal or off-the-shelf healthy cells for transplantation.

Table 1.5. Summary of studies using CRISPR/Cas9 technology to disrupt the expression of MHC-I molecules.

Target	Purpose	Type of change	Year	Reference
B2M	Effect of Antigen processing and presentation machinery on immune checkpoint inhibitors in lung cancer	Knock-out	2017	[309]
HLA-A, HLA-B, HLA-C	Cellular therapies and transplantation	Knock-out	2017	[310]
B2M	Generate a line of "universal donor" iPSCs	Knock-out	2018	[311]
B2M	Generate universal corneal epithelial cells (CEC) from human embryonic stem cells (hESC)	Knock-out	2018	[312]
ERAP2	Study HLA-B*40:02 peptidome	Knock-out	2019	[313]
B2M	Identify (HLA) antibody cross-species reactivity with swine leukocyte antigen (SLA)	Knock-out	2019	[314]
B2M HLA-A HLA-B	Generates iPSCs with Enhanced Immune Compatibility	Knock-out	2019	[315]
B2M	Induce K562 to become artificial antigen-presenting cells for ex vivo expansion of T lymphocytes	Knock-out	2019	[316]
B2M	Converts NK cells into Effectors for "Off-the-Shelf" Immunotherapy	Knock-out	2020	[317]
B2M CIITA TRA	Modification of allelic T cells for adoptive T-cell therapy	Knock-out	2020	[318]
HLA-A, HLA-B, HLA-C	Generate universal Mesenchymal stem cells (MSCs) derived from iPSCs	Knock-out	2020	[319]
B2M CIITA	Generate Retinal Pigment Epithelial Cells Derived lacking HLA class I and II	Knock-out	2020	[320]
B2M	Generate Allogeneic human umbilical MSC source for stem-cell therapy	Knock-out	2020	[321]
B2M	Generate Mesenchymal Stromal Cells with Low Immunogenicity from Human PBMC	Knock-out	2020	[322]
B2M	Generate HLA-KO Glomerular Endothelial Cells for prediction of pathogenic Non-HLA Antibodies in Kidney Transplant Recipients	Knock-out	2021	[323]

B2M	Generate B2M-KO somatic cell nuclear transfer-derived embryonic stem cell line	Knock-out	2021	[324]
B2M	Generate HLA-I depleted iPSC to overcoming the immunogenicity of allograft transplantation.	Knock-out	2021	[325]
TAP1	Investigated the ability of human colorectal cancer (CRC) to cross-present neoantigen-derived synthetic long peptides	Knock-out	2022	[326]

Most research related to CRISPR/Cas9-mediated disruption of MHC-I generally aimed to determine the role of the antigen processing pathway or peptide spectrum in particular disease states. A study by Gettingger and colleagues (2017) employed CRISPR/Cas9 to investigate the connection between HLA class I antigen processing and immune checkpoint inhibitors (ICIs) [309]. The PX459 containing Cas9 and B2M specific sgRNA were transfected into a murine lung cancer cell line (UNSCC680AJ) which was later used to generate a tumor model. While the intact tumor model responded to anti-PD-1, that of the B2M disrupted cells proceeded through PD-1 blockage which suggest its role in resistance to ICIs. The result of this study showed that disruption of HLA class I antigen processing could lead to lung cancer cells escaping ICI treatment. In Lorente's study [313], the CRISPR/Cas9 system was used to study the effect of the endoplasmic reticulum aminopeptidase 2 (ERAP2) on peptidomes of HLA-B*40:02, an ankylosing spondylitis (AS) weakly associated MHC-I molecule. ERAP2 is an enzyme that processes MHC-I peptides and favorably trims N-terminal basic residues. CRISPR/Cas9-mediated disruption of ERAP2 was performed in a human lymphoid cell line expressing HLA-B*40:02 followed by the investigation of the peptidomes, which are the complete set of peptides encoded by a particular genome, between wild-type and ERAP2-KO cells. The results showed that by disrupting ERAP2, there was a 5% qualitative change of the peptidome and

increased the proportion of N-terminal basic residues which were frequently found as ligands for HLA-B27.

Recently, induced pluripotent stem cells (iPSC) have been considered as a potential approach in the field of regenerative medicine [327]. However, allogenic transplantation of iPSC-derived differentiated cells can cause immune rejection due to HLA mismatches. HLA-I mismatches would stimulate CD8⁺ cytotoxic cells, whilst HLA-II mismatches can activate CD4⁺ helper T cells which can produce cytokines and stimulate other immune cells. On the other hand, cells whereby *B2M* had been knocked out to disrupt expression of HLA I molecules, could activate NK cells. Genome editing was employed by Xu and colleagues to generate different HLA-knock out iPSCs (or pseudo-homozygous iPSCs) to evaluate their ability to activate T and NK cells (Figure 1.16) [315]. Results showed that wide-type (WT) MHC mismatched iPSC-derived from CD43⁺ blood cells stimulated CD8⁺ T cells, whereas HLA-A-, B-, C+ cells and B2M-KO cells did not. On the other hand, a subset of NK cells only reacted to B2M-KO cells, but not to WT cells or HLA-A-, B-, C+ cells as determined by both *in vitro* T cell cytotoxicity by ⁵¹Cr release assay and *in vivo* graft survival assays on NRG mice (lacking endogenous T, B, and NK cells). Hence, the lack of HLA-A and -B was effective for evading T cell immunity and the retention of HLA-C (and intact HLA-E, -F, and -G) was sufficient for suppressing NK cell activity. However, the immune system could still be activated by CD4⁺ helper T cells after recognizing HLA class II mismatches. CIITA deficient, HLA-C+ cells, which exhibited disruption of HLA class II expression, did not stimulate the proliferation of CD4⁺ helper T cells. In Xu's study, iPSC cells with disrupted HLA-A/B, CIITA and intact HLA-C evaded both CD8⁺ and CD4⁺ T responses as well as NK cell activities. 12 common HLA-C-retained iPSC

lines were generated which were proposed to be sufficient to cover >90% of most popular ethnic groups.

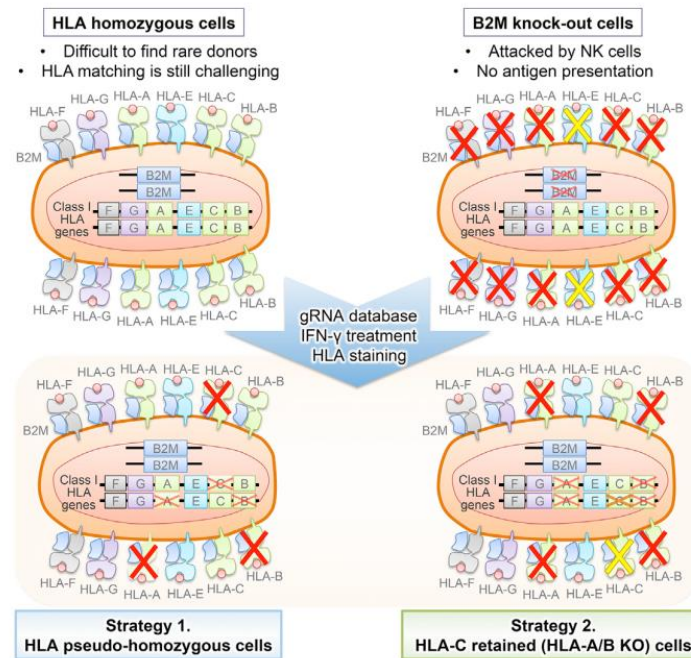


Figure 1.15. Strategies of CRISPR/Cas9-mediated disruption of HLA genes for generation of universal iPSC with enhanced immune compatibility.

Two genome-editing strategies were developed to generate immune-compatible donor iPSCs. First, HLA pseudo-homozygous iPSCs were made using CRISPR-mediated knock-out of non-specific HLA alleles. Second, HLA-C-retained iPSCs were generated by disruption of HLA-A and HLA-B alleles. Both HLA pseudo-homozygous and HLA-C-retained iPSCs were proposed to evade T cells and NK cells. Diagram was obtained from Xu *et al* [315].

1.7 Thesis rationale and objectives

Although CRISPR/Cas9-mediated disruption of MHC-I has received much attention with respect to applied research and clinical application, there are few if any studies reporting to use CRISPR/Cas9 to modify or introduce changes of interest in either HLA class I or B2M molecules. Compared to CRISPR/Cas9-mediated knock-out, knock-in procedures require more starting material and are only suitable to apply to limited cell lines. To facilitate DSB repair via the HDR pathway, the donor template must be precisely designed to contain desired nucleotide changes as well as appropriate homology arms. On the other hand, most cell lines used show preference for using the NHEJ pathway rather than HDR to correct DSBs. However, the HDR-mediated knock-in can bring greater benefits in lowering the off-target effects and introducing targeted modification. While DSB repair via NHEJ generally leads to multiple indel mutations, DSB repair via HDR pathways can introduce either in-frame or out-frame modifications depending on the design of the donor template. In the case of MHC-I, CRISPR/Cas9-mediated precise modification not only can potentially manipulate presented peptides but can also be useful to switching HLA subtypes related to diseases or causing graft rejection. In this study, we will focus on developing protocols using CRISPR/Cas9 to introduce desired modifications to both B2M and HLA class I in order to change the properties of MHC-I molecules. In addition, the off-targets caused by CRISPR/Cas9 gene-editing were also evaluated to determine the safety of this method.

To achieve this aim, the following objectives were established, including:

1. Using the CRIPSR/Cas9 system to introduce peptide and linker sequences into the B2M gene to create and characterize the novel antigen-presenting molecules.
2. Using CRISPR/Cas9 to switch a disease-associated HLA to a non-disease associated HLA subtype.
3. Using next generation sequencing to evaluate the off-target effects on MHC-I modified cell models.

CHAPTER 2
MATERIALS AND METHODS

Chapter 2 Materials and Methods

2.1 Cell culture

2.1.1 Culture Cell

Cell lines used in this thesis are listed in table 2.1. All cell culture was performed in class II safety cabinets. Dulbecco's Modified Eagle Medium (DMEM) and Roswell Park Memorial Institute (RPMI) medium was supplemented with FBS (10%-Gibco), L-glutamine (2mM) and Penicillin/Streptomycin (100 U/mL).

Newly purchased cells or frozen cells were defrosted in a 37°C water bath, resuspended in 5mls of pre-warmed growth medium, centrifuged at 300g for 3 minutes followed by resuspension in 1 ml growth medium or Phosphate-buffered saline (PBS). Cell numbers were determined following staining with the live-dead cell by trypan blue dye.

Adherent cells were seeded at a density of 5×10^4 cells/cm² and 10^5 cells/ml for suspension cells, followed by incubation at 37°C, in a 5% CO₂ humidified incubator. Cells were passaged once they reached 70-80% confluency. Adherent cells were washed briefly in 1xPBS, treated with 0.05% trypsin-ethylenediaminetetraacetic acid (EDTA) 40 µl/cm² and incubated for 3 minutes at 37°C. Following cells detachment, pre-warmed medium was added at 5x volume of 0.05% trypsin-EDTA to inactivate enzyme activity. Suspension and detached adherent cells were centrifuged at 300g for 3 minutes and seeded as described.

For freezing, cells were made to 10^6 cells/ml in Fetal Bovine Serum (FBS) (90%) and dimethyl sulfoxide (DMSO) (10%) freezing mix and 1 ml of aliquots were stored at -80°C, followed by storage at -152°C.

Table 2.1. Information of cell lines and growth medium

Full name	Type	Growth medium	Cell Type	Reference
HeLa	Cervical carcinomas cells	DMEM	Adherent	[328]
K562	Human immortalised myelogenous leukemia cell	RPMI	Suspension	[329]
HEK 293	Human embryonic kidney	DMEM	Adherent	[330]
T2	Lymphoma-derived cell line	RPMI	Suspension	[331]

2.1.2 Mammalian cell transfection

DNA Transfection

JetPRIME lipofection reagent was used to transfect all plasmids into mammalian cells. Amounts of lipofection reagent and plasmids were listed in table 2.2. Once cells reached optimal confluence (60-80%), plasmid DNA was diluted in jetPRIME buffer, followed by the appropriate volume of JetPRIME reagent. The DNA-JetPrime reagent mix was incubated 10 mins at RT, then added dropwise onto the cells and incubated at 37 °C. Media was replaced at 4 hours post-transfection and analyzed after 24 hours post-transfection.

Table 2.2. Amount of jetPRIME reagents and DNA according to the cell culture vessel per well

Culture Vessel	jetPRIME® Buffer (µl)	DNA (µg)	jetPRIME® reagent (µl)	Growth medium (ml)
24-well	50	0.5	1	0.5
12-well	75	0.8	1.6	1
6-well	200	2	4	2

CRISPR/DNA Template Transfection

CRISPRMAX Reagent Cas9 nuclease transfection protocol provided by ThermoFisher was used to deliver Cas9 RNP into mammalian cells. The amount of transfection reagent and Cas9 RNP to be used were listed in table 2.3. Cas9 RNP (Cas9 nuclease + gRNA) solution, Cas9 Plus Reagent and Opti-MEM™ medium were added and mixed (Tube 1). CRISPRMAX Reagent and Opti-MEM™ medium were mixed separately (Tube 2). Tube 1 contents were transferred to tube 2 and incubated 10 minutes at room temperature, followed by adding the transfection mix dropwise onto cells. After incubating at 37 °C for at least 1-day, transfected cells were ready for analysis.

Table 2.3. Amount of CRISPRMAX reagents, Cas9 nuclease and sgRNA according to the cell culture vessel per well

Cultured Vessel	Growth medium (mL)	Tube 1				Tube 2		Cas9 RNP/transfection reagent complex (µl)
		Opti-MEM™ medium (µl)	Cas9 nuclease (µg)	gRNA (µg)	Cas9 Plus Reagent (µl)	Opti-MEM™ medium (µl)	CRISPRMAX Reagent (µl)	
24-well	0.5	25	0.625	0.125	1.3	25	1.5	50
12-well	1	50	1.25	0.25	2.5	50	3	100
6-well	2	125	3.15	0.63	6.3	125	7.5	200

Nucleofector of Cas9 RNP and dsODN/Megamer.

For HDR, Cas9 RNP and donor templates were delivered into mammalian cells by electroporation using Amaxa Nucleofector and Cell line Nucleofector Kits (Lonza).

Nucleofection conditions of different cell lines were listed in table 2.4. Cell culture media with and without HDR enhancer were prewarmed to 37°C before transfection. Once cells reached optimal confluence they were carefully harvested (by trypsinization for adherent cell lines), spun down and counted. 1×10^6 cells were aliquoted into microcentrifuge (1.5 mL) tubes for each transfection, centrifuged 200xg for 10 minutes at room temperature and supernatants removed completely. The transfection mixture was prepared according to table 2.4. Each cell pellet was then resuspended in the transfection mixture, transferred to a transfection cuvette and electroporated according

to the transfection program suggested by the manufacturer. Transfected cells were carefully resuspended in 500 μ l of culture medium and transferred to 1.5 ml culture media containing Alt-R HDR enhancer. Cells were incubated at 37°C, 5% CO₂ in a humidified incubator for at least 24 hours prior to analysis.

Table 2.4. Nucleofection conditions and amounts of HDR enhancers, Cas9 nuclease, sgRNA and ssODN according to cell lines employed.

Cell line	Kit/Program	Transfection mixture					Alt-R HDR Enhancer (μ l)
		Cas9 nuclease 62 μ M (μ l)	gRNA 10 μ M (μ l)	PBS (μ l)	ssODN/ Megamer (μ g)	Alt-R Cas9 Electroporation Enhancer (μ l)	
HeLa	R/ A-28	0.5	3.1	1.4	1.2/2	1.2	10
K562	V/ T-16						
HEK	V/ Q-01						
EBV B	C/X-01						

2.1.3 Flow cytometry analysis

MHC-I cell surface expression was detected using different antibodies (Table 2.5). K562 and HEK cells, which express low levels of MHC-I, were treated with Interferon gamma (IFN- γ) at 1000 IU/ml and 400 IU/ml, respectively, for 3 days prior to flow cytometry analysis. Briefly, cells were harvested and centrifuged at 400xg for 3 minutes followed by resuspension in ice cold PBS/1%FBS. Approximately 2×10^5 cells were aliquoted into a 96 well (U-bottom) per antibody staining. Cells were incubated with appropriate antibody for 30 minutes at 4°C, followed by 3x washes with PBS/1%FBS, followed by incubation with secondary antibody for 30 minutes at 4°C.

After washing with PBS/1%FBS, samples were acquired on a BD FACSCanto™ II Clinical Flow Cytometry System. Fluorescence-activated cell sorting (FACS) data was then analyzed using Flowjo software (FlowJo v10) and GraphPad Prism (version Prism 9.0.0).

Table 2.5. List of antibodies used for flow-cytometry

Name	Epitope	Concentration/ Volume	Antigen species	Supplier	Catalogue No./Ref
BB7.1	HLA-B*07	20µl/ml	Mouse	Abcam	[332]
BB7.2	HLA-A*02	20µl/ml	Mouse	Biologend	343302
W6/32	HLA-A,-B,-C,- G,-E,-F + B2M	20µl/ml	Mouse	N/A	[333]
ME1 (mAb)	HLA-B27	50 µl/sample	Mouse	Dr. Simon	[334]

2.1.4 Single cell cloning

For single cell cloning, initially polyclonal cells were harvested at 60-70% confluency and counted. Cells were then serially diluted to 10^3 cells/mL in complete medium. 100 µl and 30 µl of 10^3 cells/mL were transferred to 10 mL complete medium to achieve 1 and 0.3 cells/well, respectively, and plated at 100 µl/well of a 96-well flat-bottomed plate. Plates were incubated at 37°C for 14 days and single clones were transferred to a 48-well plate for expansion. After sufficient expansion, lines were screened by

molecular techniques/flow-cytometry analysis and then further expanded for later experimentation or storage.

2.2 Molecular biology

2.2.1 Bacterial Strains

Escherichia coli (*E.coli*) bacterial strains used in this study included DH5 α , Stella and TOP10. A loop of *E. Coli* stock or 20 μ l of *E. coli* was spread on Luria Broth (LB) agar plates (no selection antibiotics) and incubated overnight at 37°C. Colonies were picked and sub-cultured in LB medium overnight at 37°C for expansion, plasmid extraction and storage.

2.2.2 Bacterial transformations

Competent cells were thawed on ice for 20-30 minutes and incubated with approximately 50ng of plasmid DNA for 30 minutes on ice. Cells were heat shocked at 42°C for 30 seconds, incubated on ice for 2 minutes followed by the addition of 100 μ l of SOC medium (ThermoFisher), and incubated at 37°C for 60 minutes. Transformed bacteria were spread onto LB agar/ampicillin (100 μ g/ml) plates and incubated overnight at 37°C.

2.2.3 Extraction of genomic DNA, RNA and plasmid extraction

Genomic DNA extraction by column. PureLink™ Genomic DNA Mini Kit was used for extracting genomic DNA. Cells (up to 5x10⁶ cells) were collected into a 15 ml conical tube and centrifuged at 300g for 3 minutes. After removing supernatant, cell pellets were resuspended in 200 μ l PBS and transferred to centrifuge tubes. Cell suspensions were mixed with 20 μ l Proteinase K and 20 μ l RNase A, and then incubated at room temperature for 2 minutes. 200 μ l PureLink® Genomic Lysis/Binding Buffer was then

added, vortexed vigorously to form a homogenous solution followed by incubation at 55°C for 10 minutes. 200 µl 100% ethanol was added to each cell lysate and mixed well by vortexing. Mixtures (~640 µl) were transferred into PureLink® Spin Columns with collection tubes, then centrifuged at 10,000g for 1 minute at room temperature. The columns were washed with 500 µl wash buffer 1 and 2, by centrifugation at 10,000g for 1 minute and centrifuged at 13,000g for 3 minutes, respectively. DNA was eluted in 50-100 µl elution buffer following incubation at room temperature for 1 minute and centrifugation at 13,000g for 1 minute. Quality and quantity of DNA samples were determined using NanoDrop One (Thermo Fisher Scientific). DNA with OD 260nm : 280nm between 1.80 and 2.00 were used for further experimentation or stored at -20°C.

Genomic DNA extraction by QuickExtract solution. QuickExtract™ DNA Extraction Solution (Lucigen) was used for extracting genomic DNA. Cells (up-to 10⁴ cells) were collected into 1.5ml centrifuge tubes and centrifuged at 300g for 3 minutes. After removing the supernatant, cell pellets were resuspended in 0.5 mL of QuickExtract Solution following by vortexing for 15 seconds. Tubes were transferred to a heat block at 65°C and incubated for 6 minutes followed by vortexing for 15 seconds. Tubes were transferred to a heat block at 98°C and incubated for 2 minutes. Extracted DNA samples were used for PCR amplification using a volume ratio 1:10 (DNA/PCR mixture). Quick-extracted DNA was stored at -20°C for up to 1 week, or at -70°C for longer periods.

RNA extraction by RNeasy Mini kit. The RNeasy extraction procedure was used to extract and purify RNA from cultured cells. Cells were grown to 70%-80% confluency in a 6-well plate and trypsinized for collection. Total RNA was extracted by following

the manufacturer's conditions. Briefly, cells (up-to 10^7 cells) were collected into 1.5 ml centrifuge tubes and centrifuged at 300g for 5 mins. After removing supernatant, cell pellets were disrupted by adding Buffer RLT (350 μ l for $< 5 \times 10^6$ cells and 500 μ l for $5 \times 10^6 - 10^7$ cells). The lysates were pipetted directly into a QIAshredder spin column followed by centrifugation at 10000g for 2 minutes and adding volume of 70% ethanol to the homogenized lysate. Samples were transferred to RNeasy spin columns following centrifugation at 8000g for 15s. After discarding the flow-through, columns were washed by respectively adding 700 μ l Buffer RW1 and 500 μ l Buffer RPE x 2 to the columns followed by centrifugation at 8000g for 15s. RNA samples were eluted in 50 μ l RNase-free water and checked for quality and concentration using a Nanodrop One. RNA samples were used directly for cDNA synthesis or stored at -20°C .

Plasmid extraction by NZY Speedy Miniprep. Single colonies were picked from selective plates, inoculated into 2-5 ml LB medium containing the appropriate antibiotic followed by incubation at 37°C for approximately 16 hours. Cells were centrifuged at 12,000g for 3 minutes and supernatants discarded. Cell pellets were resuspended by respectively adding 250 μ l of Buffer A1, 250 μ l of Buffer A2 and 300 μ l of Buffer A3 followed by centrifugation at 12,000g for 3 minutes. Supernatants were transferred to NZYTech spin columns followed by centrifugation at 12,000g for 30 seconds. 600 μ l of Buffer AS was added into each column following centrifugation at 12,000g for 1 minute. DNA plasmids were eluted by adding 50 μ l of Buffer AE following incubation for 1 minute before centrifugation at 12,000g for 1 minute. Plasmid DNA samples were used directly for further experiments or stored at -20°C

Plasmid extraction by QIAGEN Midi kit. Single colonies were collected from selective plates and inoculated in 2 ml LB medium containing the appropriate antibiotic followed

by incubation at 37°C for approximately 8 hours. All bacterial cells were transferred into 50ml LB medium containing the appropriate antibiotic following incubation at 37°C overnight. Cells were transferred to 50 ml tubes and centrifuged at 6000 x g for 15 min at 4°C. Cell pellets were resuspended by respectively adding 0.3 ml of Buffer P1, 0.3 ml of Buffer P2 and 0.3 ml of chilled Buffer P3 followed by centrifugation at 13 000 rpm for 10 min. Supernatants were transferred to equilibrated QIAGEN-tip 20 and allowed to enter the resin by gravity. QIAGEN-tip 20 were washed twice with 2 ml Buffer QC. After eluting with 0.8 ml Buffer QF, plasmid DNA was precipitated by 0.7 volumes of isopropanol following centrifugation at 15,000 x g for 30 minutes. DNA pellets were washed with 1 ml of 70% ethanol followed by centrifugation at 15,000 x g for 10 minutes. After air-drying the pellets for 10 minutes, plasmid DNA was redissolved in 500 mL of TE buffer (pH 8.0).

2.2.4 cDNA synthesis

CDNA synthesis was performed using ThermoScript cDNA synthesis kit. Briefly, 50ng total RNA was added into PCR tubes containing 500ng oligo dT (50 µM), 200ng dNTP (10 mM) Mix and dH₂O (up-to-12 µl). The mixture was incubated at 65 °C for 5 minutes and placed on ice before adding 1x cDNA synthesis buffer, 0.1µM DTT (1 µl), RNaseOUT, dH₂O and 15U ThermoScript™ RT. The mixture was incubated at 25°C for 10 min, 50°C for 60 min and 85°C for 5 min, respectively, to generate the final cDNA product. Synthesized cDNA was directly used for PCR or stored at -20°C.

2.2.5 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is the molecular method for amplifying a desired DNA sequence to extremely high copy numbers. Sources of DNA were genomic DNA extracted by PureLink™ Genomic DNA Mini Kit or QuickExtract™ DNA Extraction Solution, purified plasmid, and purified PCR amplicons. PCR primers were designed using the primer designing tool of NCBI (<https://www.ncbi.nlm.nih.gov>). Site direct mutagenesis primers were designed by using the QuikChange Primer Design Program (Agilent). Primers used for Next Generation sequencing (NGS) were designed using COSMID (<https://crispr.bme.gatech.edu/>). Sequences of all primers were listed in Appendix 5. PCR components used for master mixes are listed in table 2.6.

Table 2.6. Components of PCR

	Components	Volume (µl)
PCR (Taq Polymerase)	Forward Primer (10 µM)	0.5
	Reverse Primer (10 µM)	
	dNTP (10uM)	0.5
	10x Taq Buffer	2.5
	Taq Polymerase	0.125
	DNA (50ng/ µl)	variable
	Quick Extract solution	variable
	Nuclease-Free Water	Up to 25 µl
Mutant Strand	SDM forward primer (100ng/µl)	1.25
Synthesis	SDM reverse primer (100ng/µl)	1.25
Reaction	10x reaction buffer	2.5
	dNTP (10uM)	1
	DNA (50ng/ µl)	variable
	<i>Pfu</i> DNA polymerase (2.5 U/µl)	1

	Nuclease-Free Water	Up to 50 μ l
Q5 High-Fidelity 2X	Q5 High-Fidelity 2X Master Mix	1.25
Fidelity 2X	Forward Primer (10 μ M)	1.25
Master Mix	Reverse Primer (10 μ M)	1.25
	DNA (50ng/ μ l)	variable
	Nuclease-Free Water	Up to 25 μ l

2.2.6 Agarose gel electrophoresis

Agarose gel electrophoresis was used to visualize DNA fragments or PCR products. Agarose powder was dissolved in 1X TAE buffer (40 mM Tris-base, 20mM acetic acid, 1 mM EDTA, pH 7.5–8.0) by heating in the microwave and allowed to cool to room temperature prior to gel casting. SYBR™ Safe gel stain was used at 1 in 10000 dilution. Ethidium Bromide or SYBR™ Gold Nucleic Acid Gel Stain were diluted 10,000X in 1X TAE buffer and used for post-staining agarose gels for 15 - 30 minutes. DNA was resolved in 1X TAE buffer and run at 100V for 60 minutes. Gel images were captured using a G-BOX F3 gel doc system (Syngene) or Molecular Imager PharosFX™ Plus System (Bio-Rad).

2.2.7 PCR purification

AMPure XP Bead Purification was used to purify PCR products for *in vitro* cleavage, *in vivo* cleavage, and Restriction Fragment Length Polymorphism (RFLP) assays. AMPure XP Bead and DNA samples were mixed at a ratio of 1.8:1 and incubated for 5 minutes at room temperature. Reaction tubes/plates were placed onto a magnetic plate for 2 minutes to separate beads from the solution. Each reaction was washed twice with 200 μ l of 70% ethanol for 30 seconds at room temperature on the magnetic plate. After removing the tube/plate from the magnetic plate, DNA fragments were

resuspended in 30-40 μl of elution buffer and incubated for 2 minutes at room temperature. The reaction tube/plates were placed onto a magnetic plate for 1 minute to separate beads from the DNA fragments. Finally, the purified PCR amplicon solution was transferred to a fresh tube/plate (Figure 2.1).

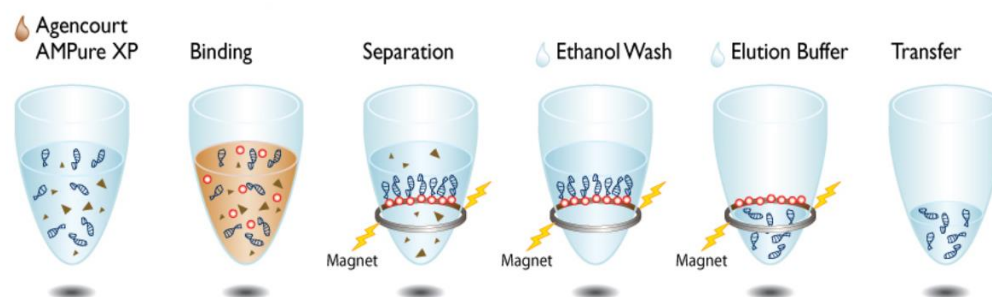


Figure 2.1. Workflow of AMPure XP Bead Purification

QIAquick Gel Extraction Kit was used to extract and purify DNA fragments (70bp to 10kb) from agarose gels and enzymatic reactions. DNA products were resolved by 1-1.5% agarose gel electrophoresis and desired DNA fragments were excised using a sharp scalpel. For each 1 volume of gel (100 mg ~ 100 μl), 3 volumes of Buffer QG were added. Gel-buffer mixes were incubated at 50°C for 10 minutes to dissolve the gel, followed by the addition of 1 volume of isopropanol and then transferred to a QIAquick spin column. The QIAquick column was centrifuged at 12000g for 1 minute and flow-through discarded. 0.75 ml Buffer PE was then added, centrifuged for 1 minute and flow-through discarded. QIAquick columns were centrifuged for an additional 1 minute to completely remove any residual buffer. Finally, DNA fragments were eluted in a fresh tube by adding 50 μl of Buffer EB and centrifuged for 1 minutes at 15000g.

2.2.8 Cell lysis

1% Nonyl phenyl-polyethylene glycol (NP-40) (US Biological) lysis buffer was made up of 150mM NaCl, 10mM Tris, pH7.5, supplemented with 10mM N-methylmaleimide (NEM) (Sigma), 1mM phenylmethylsulphonyl fluoride (PMSF) (Sigma) and 1x complete protease inhibitors (CPI) (Roche Applied Science). Cells were harvested and incubated in pre-cooled 20mM NEM (pH 7.4) for 15 minutes on ice. 100 µl of 20mM NEM was used for every 10⁶ cells. Cells were pelleted and supernatant was discarded, followed by lysis in pre-cooled lysis buffer, which was used at 100 µl per 10⁶ cells, for at least 30 minutes on ice followed by centrifugation at 13,000 g for 30 minutes. Cell lysates were transferred into pre-cooled 1.5 mL tubes. 10 µl of cell lysates were aliquoted to determine protein concentration using the Pierce™ BCA Protein Assay Kit (Thermo Scientific™). The remainder of the cell lysates were mixed with 1 volume of 2x loading buffer containing 4% sodium dodecyl sulfate (SDS), 20% glycerol, 0.1% bromophenol blue and 0.125M Tris, pH6.8 before resolving on SDS-PAGE gel or storage at -20°C. To reduce samples, lysates were incubated with 100mM dithiothreitol (DDT) followed by heating at 95°C for 5 minutes.

2.2.9 Immunoblotting

To detect high molecular weight proteins (>20 kDa), samples were resolved by 10% Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) containing 10% resolving gel (10% Acrylamide (Severn Biotech), 0.1% SDS, 0.375M Tris pH 8.8, 0.1% ammonium persulfate (APS) and 0.04% N,N,N',N'-Tetramethyl ethylenediamine (TEMED)) and a stacking gel (5% Acrylamide, 0.1% SDS 0.125M Tris pH 6.8, 0.075% APS and 0.075% TEMED). To detect low molecular weight proteins (<20 kDa), samples were resolved by 17% SDS-PAGE containing 17%

resolving gel (17% Acrylamide, 0.1% SDS 0.375M Tris pH 8.8, 0.1% APS and 0.04% TEMED) and stacking gel (as above). The PageRuler™ Prestained Protein Ladder (Thermo Scientific™) was used as the molecular weight marker to determine the size of protein bands. Equal amounts of protein (5-20 µg) were resolved by SDS-PAGE and run at 100V.

Gels were placed into a sandwich cassette and transferred onto nitrocellulose membrane (BA85, Whatmann) in 1x transfer buffer (25 mM Tris, 192 mM glycine) at 100V for 1 hour at 4°C. Following transfer, membranes were rinsed briefly in water followed by staining with Ponceau S to visualize proteins. Ponceau S stain was removed by rinsing in TBST (0.1% Tween (Sigma-Aldrich), 150 mM NaCl, 20mM Tris, pH 7.5) 3 times for 10 minutes. Membranes were blocked in 5% (w/v) skimmed milk powder in TBST (blocking buffer) at room temperature for 1 hour following incubation with appropriate primary antibody (Table 2.8), diluted in blocking buffer, overnight at 4°C. Blots were rinsed 3 times for 5 minutes with TBST following incubation with appropriate HRP conjugated secondary antibody diluted in blocking buffer for 1 hour at room temperature. Blots were rinsed 3 times for 5 minutes followed by developing using SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Scientific™). Chemiluminescent signals were captured using G: Box F3 (Syngene) and analyzed using GeneTools (version 4.3.8) to determine the band intensity of the target proteins.

Table 2.7. The list of antibodies used for immunoblotting.

Name	Epitope	Concentration	Antigen species	Supplier	Catalogue No/Ref
HC10 (mAb)	Unfolded HLA-B,-C	1:5000	Mouse	Prof. T. Elliott (Oxford University, U.K.)	[335]
pK (mAb)	V5 Tag	1:5000	Rabbit	Serotec	N/A
Anti-GAPDH (pAb)	GAPDH	1:5000	Rabbit	Abcam	ab9485
Anti-PDI	PDI	1:5000	Mouse	Abcam	ab52587
Anti-B2M (mAB)	B2M	1:5000	Mouse	Abcam	ab759

2.2.10 Restriction enzyme digestion and ligase

All restriction enzyme reactions were performed following the manufacturer's conditions. The digests were prepared as shown in table 2.8. The input DNA included plasmids or purified PCR amplicons. Incubation time was typically 1 hour at 37°C, with an extension to 3 hours if required. All experiments included a negative control comprising all reagents except the enzyme. The digested products were analyzed by using agarose gel electrophoresis and purified using QIAquick gel extraction kit if necessary.

Table 2.8. Components of restriction digest

Component	Amount
10X Restriction Enzyme	2.5 μ l
DNA	100 – 500 ng
Enzymes	0.5 μ l
Nuclease-free water	Up-to-25 μ l

2.2.11 Plasmid transformation

All constructed plasmids were selected and expanded by transforming into competent bacterial cells. Before starting the transformation process, the LB agar plates were prepared by dissolving LB agar tablets (1 tablet per 50 mL water), followed by autoclaving and cooling to 50°C prior to adding ampicillin (100 mg/ml). Approximately 12 ml of LB agar was poured into each 90 mm Petri-dish and left to solidify for 20 minutes at room temperature under aseptic conditions.

All transformations were performed by following a heat shock protocol. Competent cells were taken from -80°C and placed on ice. 100-500ng DNA was incubated with 20-50 μ l competent cells on ice for 30 minutes, followed by heat shock at 42°C for 45 seconds and incubated on ice for 2 minutes. Transformed cells were recovered by adding 500 μ l of SOC medium and incubated at 37°C for 60 minutes. All cells were spread onto LB agar plates and incubated overnight at 37°C.

2.2.10 Site-directed mutagenesis

Nucleotide changes were introduced using the QuikChange site-directed mutation (SDM) protocol. Mutagenic primers were designed using the web-based QuickChange

Primer Design Program (www.agilent.com/genomics/qcpd). The SDM PCR reaction mix included: 1× reaction buffer, 25 ng plasmid, 125 ng of each forward and reverse mutagenic primer, dNTP (10 mM) and made up to a final volume of 50 µl with double-distilled water (ddH₂O). *Pfu* DNA polymerase (2.5 U) was added to the mixture. The following thermal cycle programme was used: 95°C for 30 seconds and 16 cycles of 95°C for 30 seconds, 55°C for 1 minute and 68°C for 1 minute per Kb of plasmid length. Following amplification, each reaction was incubated with *Dpn* I restriction enzyme (10 U) for 1 hour at 37°C. 1 µl of the *Dpn* I-treated DNA was then transformed into TOP10 competent cells.

2.2.11 Generating stable mammalian expression cell lines using the Flp-in system

The pcDNATM5/FRT/V5 expression vectors containing the gene of interest (GOI) were integrated into the Flp-in host cell line HeLa-H2Z cells, (which was a kind gift provided by Dr. Keith Gould (Imperial College London) and Dr Darren Nesbeth (UCL)), containing two flippase recognition target (FRT) sites via Flp recombinase-mediated integration [336]. HeLa-H2Z cells were seeded at a density of 2x10⁴ cells/cm² in 24-well plates for 24 hours prior to transfection. HeLa.H2Z cells were co-transfected with a 9:1 ratio of pOG44: pcDNATM5/GOI/FRT/V5 by JetPrime lipofection reagent following incubation at 37°C. Cells were washed and replaced with warm fresh cDMEM 4 hours after transfection. 48 hours after transfection, cells were split and grown in cDMEM selection medium containing Hygromycin B (150 µg/ml) and/or Zeocin (150 µg/ml). Selection medium was replaced every 3-4 days until foci of cells could be identified. Selected cells were expanded and integration of pcDNATM5/GOI/FRT/V5 verified by immunoblotting with anti-V5 pK antibody.

2.2.12 *In vitro* cleavage assay

To validate the activity of Cas9 ribonucleoprotein (RNP) complexes, *in vitro* cleavage analysis of targeted dsDNA was performed. Firstly, PCR amplicons were generated by using Taq polymerase master mix and appropriate primer pairs. PCR amplicons were then purified by AMPure XP Bead Purification and checked for quality and quantity by Nanodrop. Alt-R CRISPR-Cas9 crRNA, tracrRNA (IDT) were resuspended in IDTE buffer to achieve a stock concentration of 100 μM . Working solutions were diluted in Nuclease-Free Duplex Buffer to 20 μM . Finally, sgRNA duplexes were prepared by mixing 20 μM crRNA and 20 μM tracrRNA solutions to a molar ratio of 1:1. The mixture were heated at 95°C for 5 min and then allowed to cool to room temperature for duplex formation. SgRNA solution was used immediately or stored at -20°C for up to 6 months.

RNP complexes were prepared by combining sgRNA and Alt-R SpCas9 enzyme in equimolar amounts to a final concentration of RNP complex of 1 μM in 1x PBS. *In vitro* (IVT) cleavage reactions were performed by mixing DNA substrates and RNP complexes in a 1:10 molar ratio. Thus, each IVT cleavage reaction contained 1 μl of Cas9 RNP (1 μM), 1 μl of DNA substrate (100 nM), 1 μl of 1X Cas9 Nuclease Reaction Buffer (200 mM HEPES, 1 M NaCl, 50 mM MgCl_2 , 1 mM EDTA) and 7 μl of Nuclease-Free Water. Digestion reactions were incubated at 37°C for 60 minutes before adding 20 μg Proteinase K and incubated at 56°C for 10 min. Cleaved products were visualized by agarose gel electrophoresis.

2.2.13 *In vivo* cleavage assay

Following the transfection of cell lines with RNP-gRNA complexes, *in vivo* cleavage assays were performed. Firstly, sgRNA and RNP complexes were prepared as described previously. RNP complexes were delivered into target cells using CRISPRMAX Reagent Cas9 nuclease transfection following Thermal Fishers conditions. After at least one day post-transfection, gene-edited cells were harvested for DNA extraction (section 2.2.1). Genomic DNA was used as a template for PCR reactions using primers amplifying the target site with the size of PCR amplicons approximately 800 bp. PCR products were purified by AMPure Bead and eluted in nuclease-free water (minimum 16 μ l solution and 100 ng DNA per reaction). To form final heteroduplexes, 2 μ l of T7EI Reaction Buffer (10X) was added and the mixtures were incubated using the following thermal settings:

Step	Temperature ($^{\circ}$ C)	Time
Denature	95	10 minutes
Ramp 1	95-85	Ramp rate – 2 $^{\circ}$ C/sec
Ramp 2	85-25	Ramp rate – 0.3 $^{\circ}$ C/sec

After completing the thermal reaction, 2 μ L of T7 endonuclease I (1 U/ μ L) was added to each tube and T7EI reaction was incubated at 37 $^{\circ}$ C for 60 minutes. Final reaction products were resolved by agarose gel electrophoresis to differentiate the original PCR amplicons and cleaved products.

2.2.14 Next-Generation Sequencing

Next-generation sequencing (NGS) was used to detect targeted genome modifications or indel mutations. Firstly, NGS primers were designed to generate short amplicons (200-250bp) of the target sites (Appendix 5). Amplicons were generated by using Q5 High-Fidelity DNA Polymerase (NEB) and genomic DNA extracted from Cas9 RNP-targeted cells as templates. 5 µl of PCR products were separated by 1.5% (wt/vol) agarose gel to check for single-band products. Single-band amplicons were purified using AMPure XP Bead Purification. NGS sample labelling and preparation were kindly supported by Dr. Andrew Nelson and Dr. Darren Smith (Northumbria University). Sequencing-ready libraries with unique bar codes were generated using Nextara XT DNA sample preparation kit (Illumina) according to the manufacturer's protocol. Labelled NGS samples were sequenced on the Illumina Miseq at 100,000 reads. HDR and indel analysis were performed by comparing NGS results to reference genome sequences with read alignment programs Geneious (<http://www.geneious.com/>).

2.3 Bioinformatics and statistical analysis

2.2.1 Plasmid map and sequence retrieval

The map of pCR3 vector and pcDNA5/FRT were retrieved from Addgene and Life technology (Appendix 3, Supplementary Data 1). The sequence of B2M was retrieved from *National Center for Biotechnology Information* (NCBI) by following genomic sequence ID 567, and mRNA sequence ID NM_004048.4. The full spike (S) surface glycoprotein and nucleocapsid (N) protein sequences were obtained for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Protein sequence data was retrieved from NCBI by following genomic sequence ID NC_45512.2 (S) and NC_004718.3 (N) in FASTA format (Appendix 3, Supplementary Data 2).

2.3.2 Data analysis and graphic illustrations

The statistical analysis in this study was performed using one-way or two-way ANOVA on GraphPad Prism (version 9.3.0) unless otherwise state. Graphic illustrations were drawn using Powerpoint, Paint, ImageJ and Biorender (<https://biorender.com/>).

CHAPTER 3

**RESCUE OF MHC-I EXPRESSION BY PEPTIDE-
LINKED B2M**

Chapter 3 Rescue of MHC-I Expression By Peptide-linked B2M

3.1 Chapter aims & objectives

The aim of this chapter was to generate novel peptide linked molecules that can facilitate the stable expression of MHC-I molecules, which can then be generated using gene-editing technology. Based on the concept of the construction of MHC-I SCT molecules, a novel fusion molecule of peptide and B2M has been developed and referred to as the **Peptide-linked B2M (PLB)** molecule. PLB molecules were developed in which the β 2m light chain and a presented peptide were joined together via a flexible linker sequence. PLB proteins were expected to be expressed and delivered to the ER where they would assemble with endogenous specific HLA class I heavy chains. If the MHC-I complexes containing PLB were successfully folded, these molecules would be transported to the plasma membrane to present the desired peptide.

To achieve the aim, several objectives were established, including:

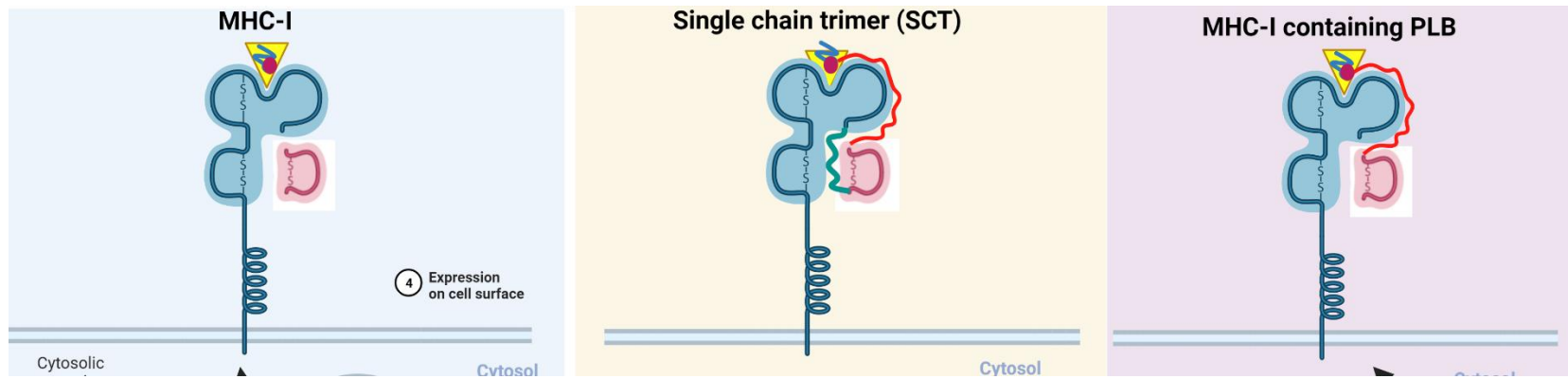
1. Designing of peptides of SARS-CoV-2 origin which were capable of presentation by HLA-A*02.
2. Establishment of DNA plasmid vectors containing B2M cDNA and PLBs with different peptide sequences.
3. Evaluation of PLB expression in a B2M-deficient HEK cell line.

3.2 Results

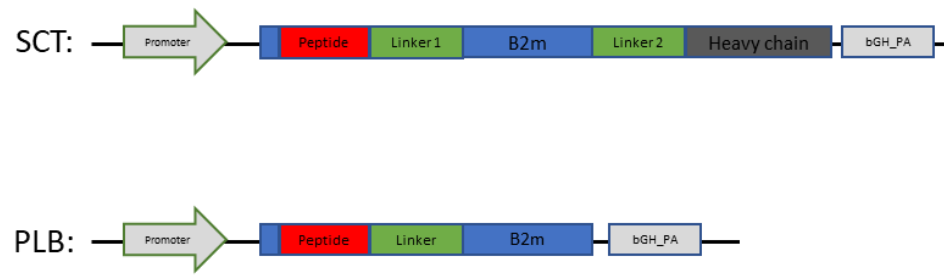
3.3.1 Strategy of Peptide-linker-B2M plasmid construction

Peptide-linker-B2M plasmids (pPLB) were constructed based on the structure of single chain trimeric (SCT) molecules in which the three components of the MHC-I complex are physically linked together (Figure 3.1A). The amino (N)-terminus of the SCT contains the peptide linked to the N-terminus of B2M by a flexible linker and the C-terminus of B2M is linked to the N-terminus of MHC-I heavy chain by a second flexible linker (Figure 3.1B) [189]. Similarly, pPLB was developed to contain the desired peptide at the N-terminus, followed by a flexible linker, with B2M at the C-terminus. The first 60 nucleotides of B2M contain the signal sequence which directs B2M translocation via the ER. Therefore, the aim was to locate the peptide of the pPLB construct between the signal sequence and the rest of the B2M sequence. By facilitating the signal sequence of B2M, the mRNA of pPLB could be translated and translocated into the ER. Therefore, the first aim of this section was to generate the pCR3.1-*B2M* plasmid by cloning the B2M cDNA into the pCR3.1 vector, followed by insertion of peptide and linker sequences into pCR3.1-*B2M* plasmid to generate pCR3-peptide-linker-*B2M* where the peptide-linker is located between Ala¹¹ and Ile¹² within the B2M protein sequence (Figure 3.1C). Both pCR3-*B2M* and pCR3-peptide-linker-*B2M* plasmids consist of the mature human *B2M* sequence but pCR3-peptide-linker-*B2M* plasmids also contain the specified peptides and linkers as stated in table 3.5.

A



A



B

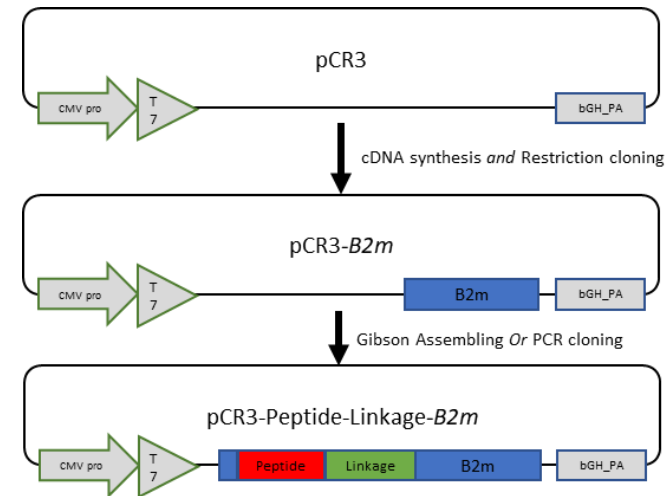


Figure 3.1. Construction of single chain trimeric molecule and peptide linked B2M molecules.

(A-B) Schematic outlining the SCT and PLB constructs in comparison to a native MHC-I-peptide complex. The SCT consists of hu*B2M* signal sequence – peptide – linker 1-- *B2M* – linker 2 – HLA class I heavy chain. The PLB exhibits a similar construction to the SCT but without linker 2 and HLA class I heavy chain. I) Cloning strategy of pCR3-*B2M* and pCR3-peptide-linker-*B2M* plasmids. Firstly, cDNA of *B2M* is cloned into empty pCR3 plasmid by restriction enzyme cloning, followed by peptide-linker insertion into pCR3-*B2M* (between Ala¹¹ and Ile¹² of *B2M* residue) using either Gibson assembly or PCR cloning.

My first aim was to clone B2M as this would form the backbone of the peptide-linker sequences I proposed to use. I proceeded to generate a *pCR3.1-B2M* plasmid, which will be the backbone for *pCR3.1-peptide-linker-B2M* plasmid and would be used as a positive control for later experiments. B2M cDNA was cloned into pCR3.1 (Invitrogen) using bidirectional cloning. cDNA was synthesized from total RNA extracted from HEK cells using ThermoScript™ RT reagents. B2M cDNA was PCR amplified using F1.B2McDNA and R1.B2McDNA cloning primers (Appendix 5) which inserted BamHI and EcoRI restriction enzyme sites located 5' and 3', respectively, to the B2M sequence. A temperature gradient PCR was performed to determine optimal PCR conditions for B2M amplification (Figure 3.2A). The gradient PCR product demonstrated the primer pair generated a specific product between the annealing temperature range 52.7°C to 62°C with little if any detectable non-specific PCR amplicons.

Purified PCR amplicons and *pCR3.1* plasmid were double digested with BamHI and EcoRI before isolating the desired fragments followed by ligation with T4 DNA ligase. Bacterial colony screening by PCR using F-CMV and R1.B2M.cDNA primers revealed 2 in 4 colonies were positive for B2M cDNA (Figure 3.2B). Sanger sequencing, followed by BLAST sequence alignment showed that clone number 2 contained the intact *B2M* gene sequence (reference gene - NM_004048.4), while clone number 1 was found to contain a missense mutation (6.V>A) with exon 1 of *B2M* gene (Figure 3.2C and 3.2D). Therefore, clone 2 was used throughout the thesis.

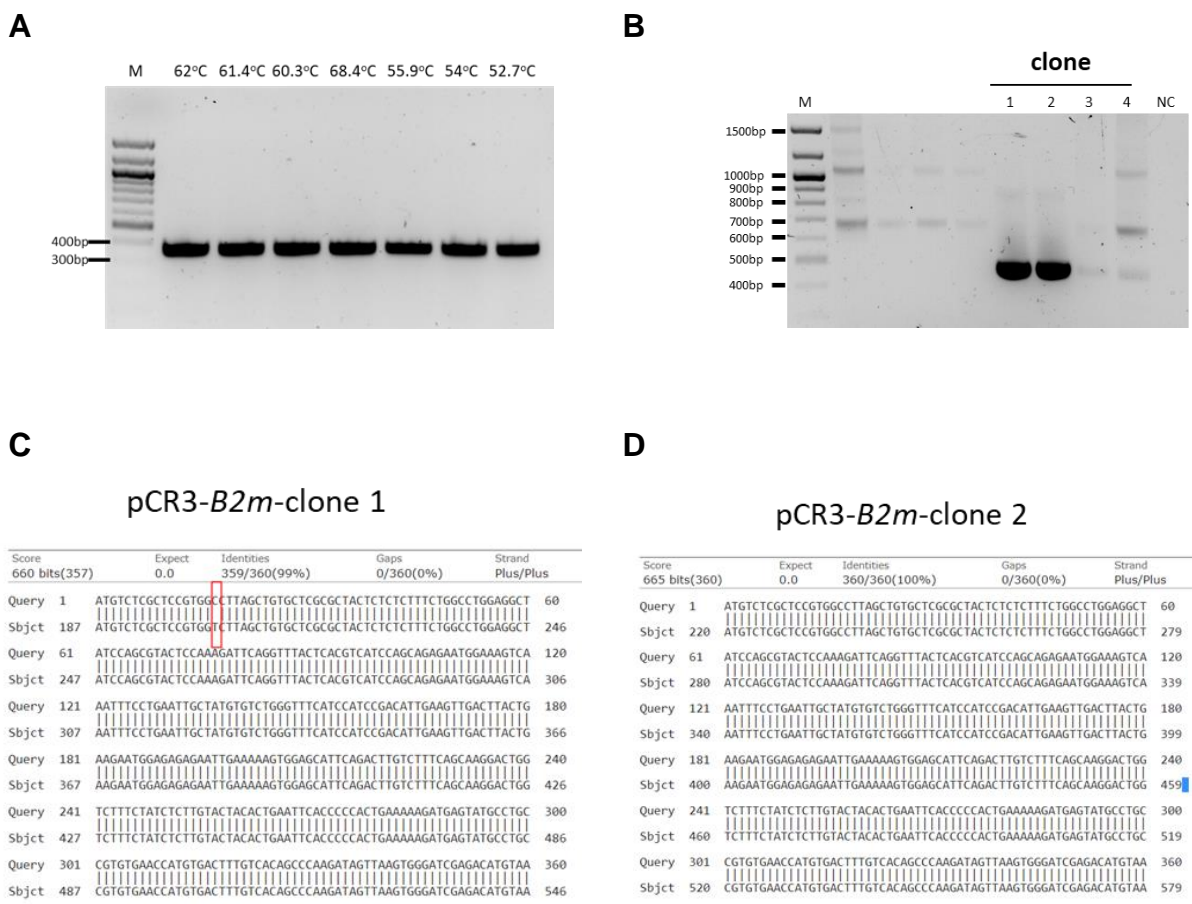


Figure 3.2. Cloning of B2M cDNA into pCR3.1 plasmid.

(A) cDNA amplicon with the size of 383 bp consisting of 360 bp B2M sequence and 2 restriction enzyme sites (EcoR1 and BamH1) inserted into the 5' and 3' flanking ends. (B) Clones containing desired pCR3.1- B2M were identified by PCR colony screening using primer pair targeting CMV promoter (F-CMV) and 5' end of B2M cDNA sequences (R1.B2M.cDNA). A PCR amplicon of 543bp was detected for clones 1 and 2 indicating successful cloning of B2m. (C, D) BLAST sequence alignment results of retrieved wild type B2M cDNA sequences and B2M amplicons from pCR3.1- B2M plasmid extracted from 2 screened clones. Nucleotide mismatch is highlighted in the red box.

3.2.2 Selection of HLA-A*02:01 specified peptides

Before constructing the pPLB plasmid, peptide and linker sequences were to be identified for testing the ability of these constructs to express peptide-B2m linked

complexes. The *HLA-A*0201* subtype was chosen as the HLA of choice due its high prevalence in all ethnic populations [337]. The selection of peptide and linker sequences was based on previous studies of SCT molecules and *in silico* analysis of pathogen-derived peptides specific for *HLA-A*02*. A literature review of SCT constructs found that there were 8 *HLA-A2* specified peptides have been used to generate SCT molecules (Table 3.1). Among them, there was the human papillomavirus HPV-E6_{29–38} (TIHDIILECV) which is one of the CD8 T-cell dominant responses to HPV infection and reported by many different investigators. The two most common linkers used to fuse peptide and B2M were G₆(SGG)₃ and (GGGGS)₃.

Table 3.1. Summary of HLA-A*02:01 specified peptides and linkers used to develop SCT molecules and stimulate CD8+ cytotoxic T cells.

Peptide sequence	Linker sequence	Origin of peptide	Epitope name	Ref
VLHDDLLEA	G6(SGG)3	HA-1 minor histocompatibility antigens	HA-1 mHag	[101]
YIGEVLVSV	G6(SGG)3	HA-2 minor histocompatibility antigens	HA-2 mHag	[101]
GILGFVFTL	(GGGGS)3	Influenza A virus	FMP58–66	[192]
TIHDIILECV	(GGGGS)3	Human papillomavirus type 16	HPV-E6(29–38)	[192, 338-340]
LLFGYPVYV	(GGGGS)3	Human T-cell leukemia virus type I	HTLV-tax11–19	[192]
ILKEPVHGV	(GGGGS)3	Human immunodeficiency virus 1	HIV-pol476–484	[192]
SVGGVFTSV	(GGGGS)4	West Nile virus	SVG9	[189]
LIYDSSLCDL	(GGGGS)4	Vaccinia virus	Mammoglobin A epitope	[189]

During the course of this study, the COVID-19 pandemic interrupted this project, however, no study had employed SCT technology to present SARS-CoV-2 peptides as a DNA vaccine. Therefore, additionally, I attempted to identify potential SARS-CoV-2-derived peptides that could be employed in the generation of PLBs.

I therefore applied a peptide prediction algorithm NetMHCpan-4.1 to predict HLA-A*02:01 binding peptides generated from the SARS-CoV-2 Spike (S) and Nucleocapsid (N) proteins. The criteria for selecting peptides included: (i) 9-mer peptides; (ii) only those specified for HLA-A*02:01; (iii) thresholds for strong and weak binders which are defined as having % eluted ligand rank <0.5 and <2, respectively [341]. By using the above criteria, the result of NetMHCpan analysis provided 1264 and 411 peptides from the S and N proteins, respectively. Based on the %Rank of the predicted binding score, the top 10 strongest binding peptides of each target were listed in table 3.2.

Table 3.2. Predicted specified HLA-A*02:01 binding peptides from S and N proteins of SARS-CoV-2.

Peptide analysis was performed by the IEDB analysis resource NetMHCpan (ver. 4.1) tool (top 10 for each protein). All peptides were 9-mers and specified for HLA-A*02:01. The score is a raw prediction score and % rank is the rank of predicted binding score compared to a set of random natural peptides.

Protein	length	peptide	score	% EL rank
S	9	YLQPRTFLL	0.971198	0.02
S	9	VLNDILSRL	0.938498	0.03
S	9	TLDSKTQSL	0.914998	0.03
S	9	RLQSLQTYV	0.87376	0.05

S	9	RLDKVEAEV	0.825045	0.06
S	9	LLFNKVTLA	0.803506	0.08
S	9	HLMSFPQSA	0.798454	0.08
S	9	VVFLHVITYV	0.74167	0.11
S	9	RLNEVAKNL	0.652653	0.16
S	9	FIAGLIAIV	0.641405	0.17
N	9	LLDRLNQL	0.955818	0.02
N	9	KLDDKDPNF	0.551611	0.23
N	9	GMSRIGMEV	0.422183	0.34
N	9	LQLPQGTTL	0.197071	0.8
N	9	DLDDFSKQL	0.190521	0.83
N	9	AQFAPSASA	0.176071	0.91
N	9	RLNQLESKM	0.126287	1.2
N	9	ILLNKHIDA	0.125549	1.2
N	9	TTLPKGIFYA	0.095317	1.4
N	9	RTATKAYNV	0.080514	1.6

Based on the literature research and *in silico* peptide analysis, I decided to use 3 different HLA-A2 specified peptides including HPV E6_{29–38}, SARS-CoV-2 N_{325–333}, and SARS-CoV-2 S_{269–277}. While HPV E6_{29–38} is a most widely used peptide with respect to experimentation relating to SCT responses, SARS-CoV-2 S_{269–277} has been predicted to exhibit the strongest binding affinity to HLA-A*02:01.

Table 3.3. Summary of selected HLA-A*02:01 specified peptides and linker for designing construct of Peptide-linked B2M (PLB)

Peptide sequence	Linker sequence	Origin of peptide	Location	PLB name
TIHDIILECV	(GGGGS)3	HPV type 16	E6 protein	E6 ₂₉₋₃₈ /(GGGGS)3
GMSRIGMEV	(GGGGS)3	SARS-CoV-2	Nucleoprotein	N ₃₂₅₋₃₃₃ /(GGGGS)3
YLQPRTFLL	(GGGGS)3	SARS-CoV-2	Spike protein	S ₂₆₉₋₂₇₇ /(GGGGS)4
YLQPRTFLL	G6(SGG)3	SARS-CoV-2	Spike protein	S ₂₆₉₋₂₇₇ /G6(SGG)3

3.3.3 Generation of pCR3.1-peptide-linker-B2M plasmids

Following selection of the desired peptides and linkers, I used the Gibson Assembly (GA) strategy illustrated in Figure 3.3A, to generate the PLB constructs using the sequences mentioned in Table 3.3. The plasmid was linearized by PCR (5130 bp) and assembled with a double-stranded DNA fragment (gBlock), containing 5' and 3' homology arms along with the peptide and linker sequences using Gibson Assembly (GA). Sequences of gBlock DNA fragments used throughout the thesis are listed in Supplementary data 2. The GA product was transformed into *E. coli* and colonies screened by PCR to identify desired clones. PCR screening was performed using primer F-CMV and R1.B2M.cDNA which would generate a 543 bp amplicon signifying only the B2M sequence or a 615-618 bp fragment containing the peptide-linker-B2M-sequence. Electrophoresis results (Figure 3.3C) indicated that 1 in 5 screened colonies (lane 3) potentially contained plasmid *pCR3.1-E6₂₉₋₃₈-(GGGGS)3-B2M*. The sequence of *pCR3.1-E6₂₉₋₃₈-(GGGGS)3-B2M* plasmid was confirmed by Sanger sequencing (Figure 3.4, top image). Electrophoresis results (Figure 3.3D) indicated 6 in 8 screened colonies (lane 1-5 and 7-8) potentially contained *pCR3.1-N₃₂₅₋₃₃₃-*

(GGGS)₃-B2M plasmid. Sanger sequencing of 3 in 6 clones showed that 2 plasmids contained the desired sequence (Figure 3.3E) with 1 clone exhibiting a mutation (Data not shown).

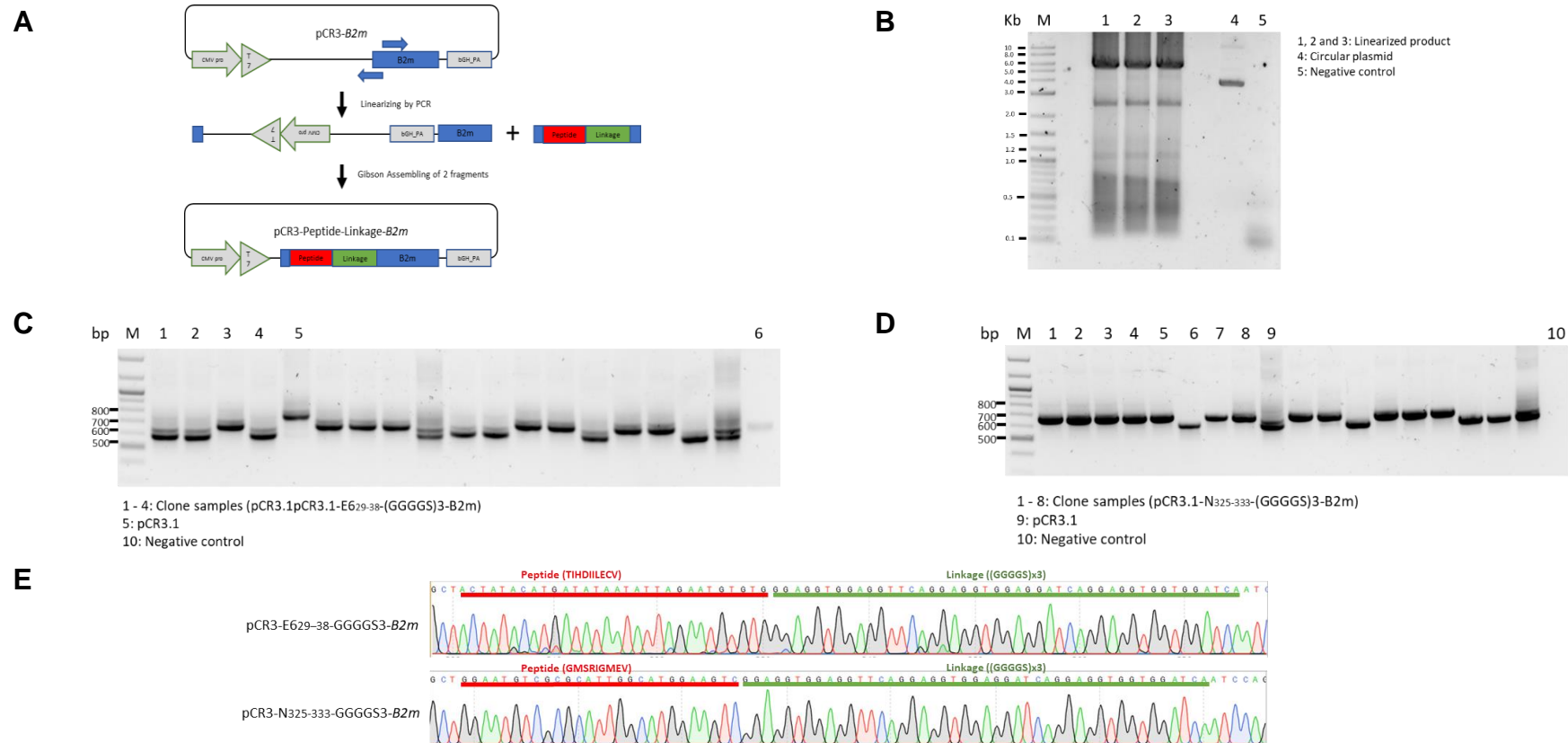


Figure 3.3. Generation of *pCR3.1-E629-38-(GGGGS)3-B2M* and *pCR3.1-N325-333-(GGGGS)3-B2M*.

(A) Strategy of inserting peptide and linker into B2M backbone sequence. (B) Electrophoresis result of linearizing *pCR3.1-B2M* by PCR reaction using B2MPI-Antisense primer and B2MPI-Sense primer. The linearized plasmid (lane 1-3) has the size of approximately 5130 bp.(C) Colony PCR screening using amplification of *-E629-38-(GGGGS)3-B2M* to identify clones containing desired inserts. PCR product

of *pCR3.1-B2M* and *pCR3.1-peptide-linker-B2M* plasmids had the size of approximately 530 and 600 bp, respectively. Therefore, Lane 3 contains the PCR product of the desired sequence. (D) Colony PCR screening using amplification of the $-N_{325-333}-(GGGGS)_3-B2M$ sequence to identify clones containing desired inserts. Lane 1-5,7 and 8 are the PCR product of the desired size containing $N_{325-333}-(GGGGS)_3-B2M$. (E) Sanger sequencing results confirm the DNA sequence of $E_{629-38}-(GGGGS)_3-B2M$ and $N_{325-333}-(GGGGS)_3-B2M$ using F-CMV primer. Data are presented of one experiment.

In order to simplify the GA cloning approach used to generate PLB plasmids, I developed the new method to clone the peptide just by including the peptide sequence within the primers used to linearize the vector alongside using the GA approach. Following this idea, I proceeded to generate *pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* plasmid as illustrated in Figure 3.4A. Using the pre-made *pCR3.1-peptide-linker-B2M* plasmid and primers (sense-GA.A2/S₂₆₈₋₂₇₇ and anti-sense- GA.A2/S₂₆₈₋₂₇₇, Appendix 5) consisting of the homologous regions with signal sequence and linker at the 5' end and the desired peptide sequence at the 3' end. The size of each primer was 60 bp. The majority of PCR products of *pCR3.1-peptide-linkage-B2M* plasmid and these primers were the linearized *pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* plasmids. The linearized plasmid has the size of approximately 5157 bp (lane 7, Figure 3.4B). Linearized *pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* plasmid were assembled by using NEBuilder HiFi DNA Assembly Master Mix. The GA product was transformed into *E coli* and screened by PCR to detect the clone containing the desired plasmid.

Simultaneously, I generated the *pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M* plasmid using the GA method with a gBlock DNA fragment by following the original method. Electrophoresis results of PCR screening reactions (Figure 3.4B) indicated that 7 in 7 screened clones (lane 1-7) potentially contain *pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* plasmids and 5 in 7 screened clones (lane 9 and 11-14) potentially contain *pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M* plasmid (Figure 3.4C). Three of each potential clones were expanded, and plasmid DNA was extracted for confirming the sequence by Sanger sequencing. Sanger sequencing results showed that all clones contained the desired *S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* and *S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M* sequence (figure 3.4D).

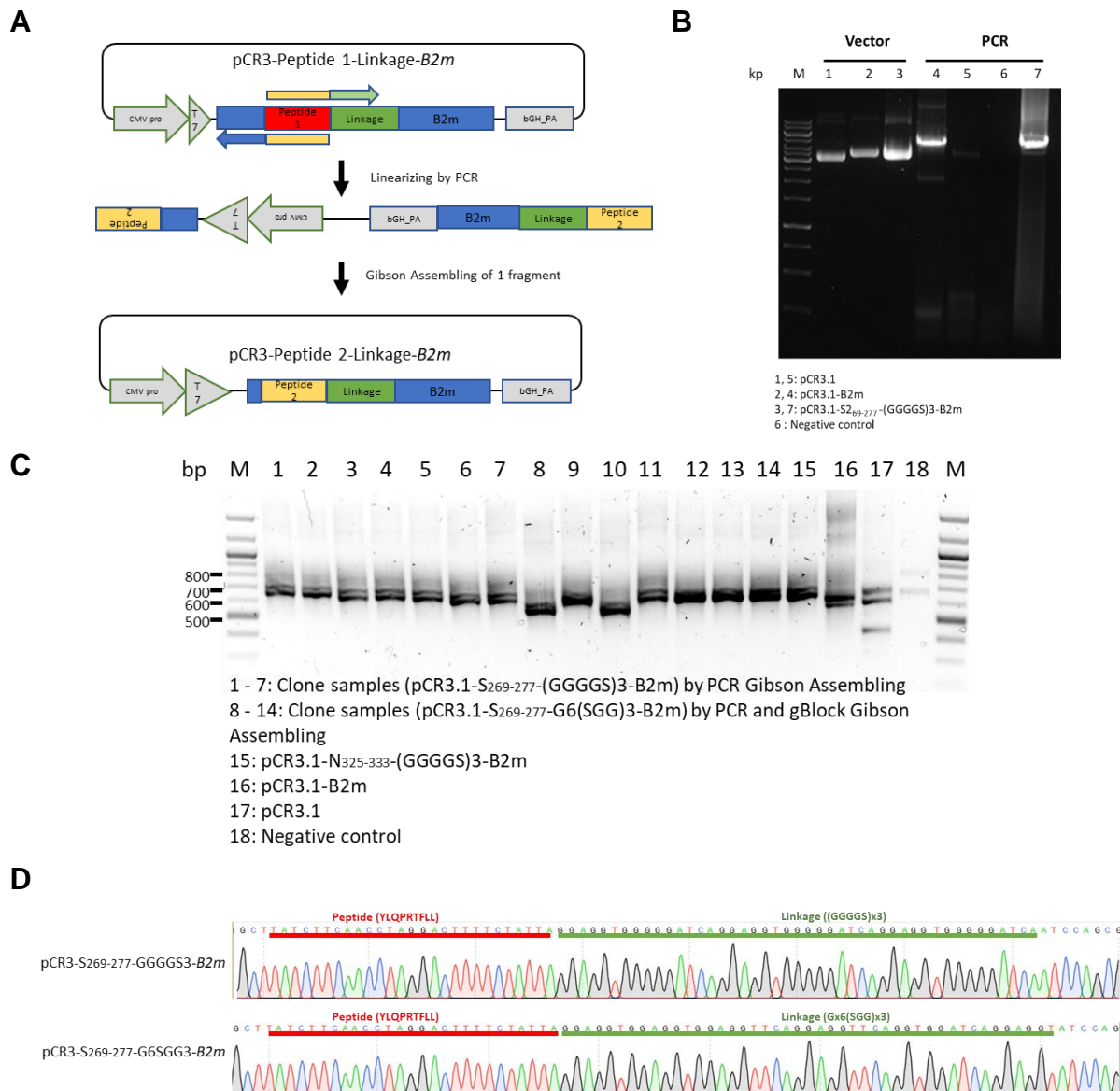


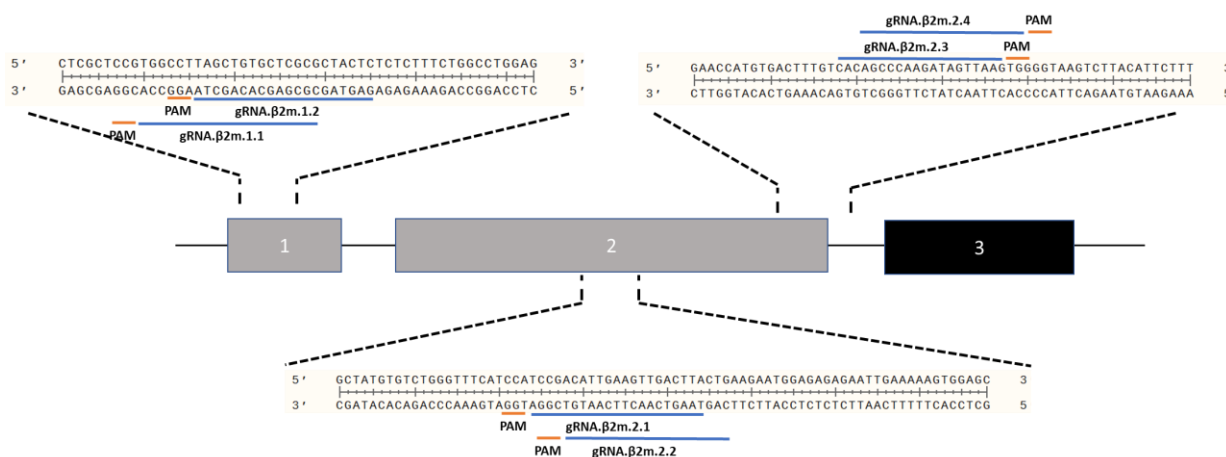
Figure 3.4. Generation of *pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* and *pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M* containing plasmids.

(A) Strategy of inserting SARS-CoV-2 S₂₆₉₋₂₇₇ peptide into *pCR3.1-peptide-(GGGGS)₃-B2M* plasmid using GA assembly with one DNA fragment. (B) Electrophoresis result of linearizing *pCR3.1-peptide-(GGGGS)₃-B2M* plasmid using sense-GA.A2/S₂₆₈₋₂₇₇ and anti-sense- GA.A2/S₂₆₈₋₂₇₇ primers. The linearized *pCR3.1-peptide-(GGGGS)₃-B2M* plasmid (lane 7) has the size of 5157 bp. (C) Colony PCR screening to identify clones containing desired inserts. The PCR product of *pCR3.1-peptide-linker-B2M* plasmids had the size of approximately 600 bp. (D) Sanger sequencing results confirm the DNA

sequence of *pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M* and *pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M* sequences matched the design construct. Data are presented of one experiment.

3.3.4 Generating a test B2M Negative HEK cell line

In order to test the expression and ability of peptide-linker-*B2M* sequences to rescue MHC-I expression I wanted to generate a B2M negative cell line. I decided to use the Cas9 ribonucleoproteins (RNPs) consisting of Cas9 protein and single-guide RNA (sgRNA) to disrupt expression of the B2M gene. Therefore, different sgRNAs were designed to target different sites of B2M by using the web based CRISPR tools called Benchling. The criteria for choosing sgRNAs were: (i) appropriate to use with *spCas9*, (ii) an optimal length of 20 nucleotides; (iii) close to the target sequence sites which included exon 1 (Chr15:25003759 and Chr15:25003765), middle of exon 2 (Chr15:25007713 and Chr15:25007716) and end of exon 2 (Chr15:25007886 and Chr15:25007888); and (iv) high score for off-target activity (>60) and on-target activity (>40). Whilst the on-target score indicates the cleavage efficiency of Cas9 RNP, that of the off-target predicts the inverse probability of Cas9 off-target binding. Based on these criteria, six different sgRNA have been identified and described in Figure 3.5. All sgRNAs were synthesized in the form of crRNA:tracrRNA duplexes (IDT).



Name	Strand	Sequence	PAM	On-target score	Off-target score	Cleavage location
gRNA.β2m.1.1	-	cgcgagcacagctaaggcca	cgg	59.2	40.5	Chr15:25003759
gRNA.β2m.1.2	-	gagtagcgcgagcacagcta	agg	48.1	47	Chr15:25003765
gRNA.β2m.2.1	-	aagtcaacttcaatgtcgga	tgg	58.2*	46.6	Chr15:25007713
gRNA.β2m.2.2	-	cagtaagtcaacttcaatgt	cgg	59.7*	39.6	Chr15:25007716
gRNA.β2m.2.3	+	cacagcccaagatagttaag	tgg	56.7*	41.4	Chr15:25007886
gRNA.β2m.2.4	+	cagcccaagatagttaagt	ggg	71.8*	41.9	Chr15:25007888

Figure 3.5. Selection of sgRNAs targeting *B2M*

Diagram (top) describes the genomic *B2M* containing 3 exons. Six different crRNAs (blue line) were designed to guide Cas9 nuclease targeting the *B2M* gene at different locations including: exon 1, middle of exon 2 and end of exon 2. The PAM sequences (NGG) specified for *spCas9* are underlined with orange. The bottom table provides details of sgRNAs designed to target the *B2M* allele.

In order to evaluate the function of the sgRNAs, two primer pairs were designed to amplify different regions within the genomic *B2M* gene. While F1.*B2M*.exon1 and R1.*B2M*.exon1 were designed to generate a 762 bp amplicon containing exon 1 of *B2M*, F3.*B2M*.exon2 and R3.*B2M*.exon2 were designed to generate a 769 bp amplicon containing exon 2 of *B2M* (Appendix 5). The function of the designed sgRNA were evaluated via an *in vitro* (IVT) cleavage assay. In this assay, equal amounts (200 µg) of naked DNA amplicons covering the target sites were treated with the RNP

complex consisting of the Cas9 nuclease and sgRNAs, followed by gel electrophoresis. If the Cas9 RNP can recognize the target site via the interaction between the crRNA and target sequence, the Cas9 protein will induce DSBs and cleave the PCR amplicons into 2 bands of different lengths, which can be visualized by 1.2% (w/v) agarose electrophoresis. The human Hypoxanthine-guanine Phospho Ribosyl Transferase (*HPRT*) sgRNA (IDT) was used as a negative control for the cleavage of B2M amplicon. Electrophoresis results (Figure 3.6) indicates that the HPRT sgRNA did not guide cleavage of either exon 1 or exon 2 *B2M* amplicons (lane 3 and 8). Cleaved bands were observed in all lanes loaded with amplicons which were treated with Cas9 RNPs targeting exon 1 (lane 1 and 2) and exon 2 of B2M (lane 4-7). The cleaved bands generated by sgRNA.B2M.1.1 (lane 1) was observed to be less intense than that of other sgRNAs, suggesting lower efficiency of cleavage directed by this sgRNA.

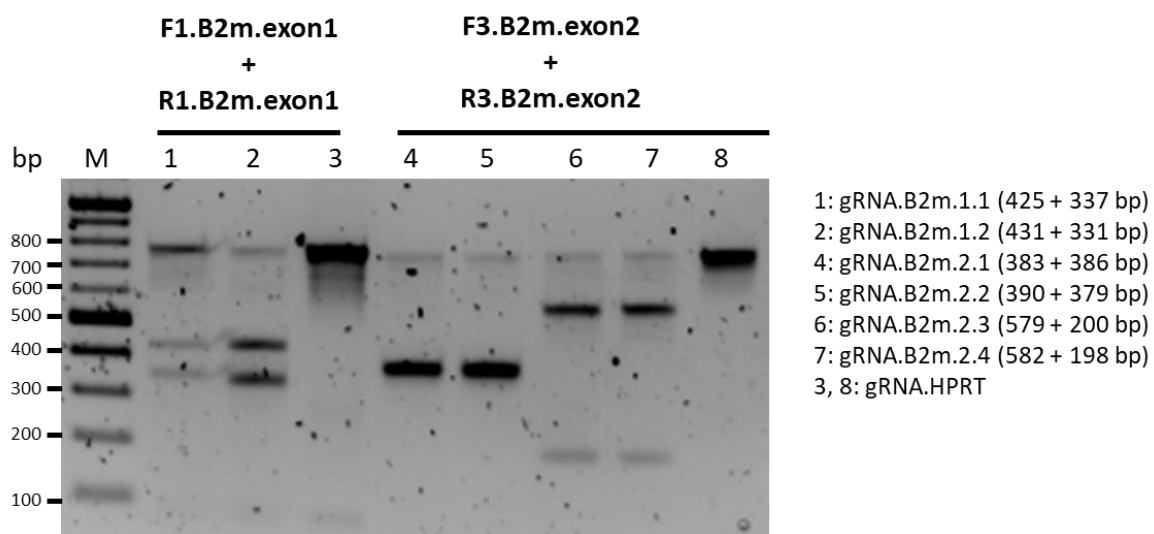


Figure 3.6. *In vitro* cleavage assay of Cas9 RNPs targeting *B2M*.

Electrophoresis results indicate cleavage of PCR B2M amplicon following incubation with Cas9 protein and sgRNAs targeting exon 1 (lane 1,2) and exon 2 of *B2M* (lane 4-7) and the control *HPRT* amplicon (lane 3,8). Amplicons of B2M exon 1 and exon 2 have the length of 762 bp and 769 bp, respectively. The annotation (right) indicates the sgRNA used for each lane and corresponding expected cleavage bands for each sgRNA. Data are presented of one experiment.

Cas9 RNPs targeting *B2M* were then used to generate the B2M-deficient cell line. HEK cells were chosen as the target cell line due to its expression of HLA-A*02:01 which is required for binding the designed peptides. Equal amounts of RNP complexes were transfected into HEK cells. Cas9 RNP targeting the *HPRT* gene was also transfected as a negative control. 24hrs after transfection, one third of the transfectants were harvested and genomic DNA (gDNA) extracted for evaluating the cleavage efficiency of Cas9 RNP targeting *B2m* by *in vivo* (IVV) cleavage assay. Genomic DNA was similarly extracted from HEK cells transfected with Cas9 RNPs targeting exon 1 and exon 2 of *B2M* and amplicons generated using F1.B2M.exon1/R1.B2M.exon1 and F3.B2M.exon1/R3.B2M.exon1 primer pairs, respectively. The PCR amplicons were denatured and reannealed to allow the formation of heteroduplexes between mutated DNA and wt DNA followed by incubation with T7 endonuclease I. CRISPR-mediated indel mutations were then detected based on the gel electrophoresis result of T7EI-treated amplicons. IVV cleavage results (Figure 3.7) of HEK cells untreated (lane 4 and 10) or treated with RNP targeting *HPRT* (lane 3 and 9) showed no cleavage. Cleavage products were observed in lanes corresponding to HEK cells transfected with Cas9 RNPs targeting B2M (lane 1, 2 and 5-8).

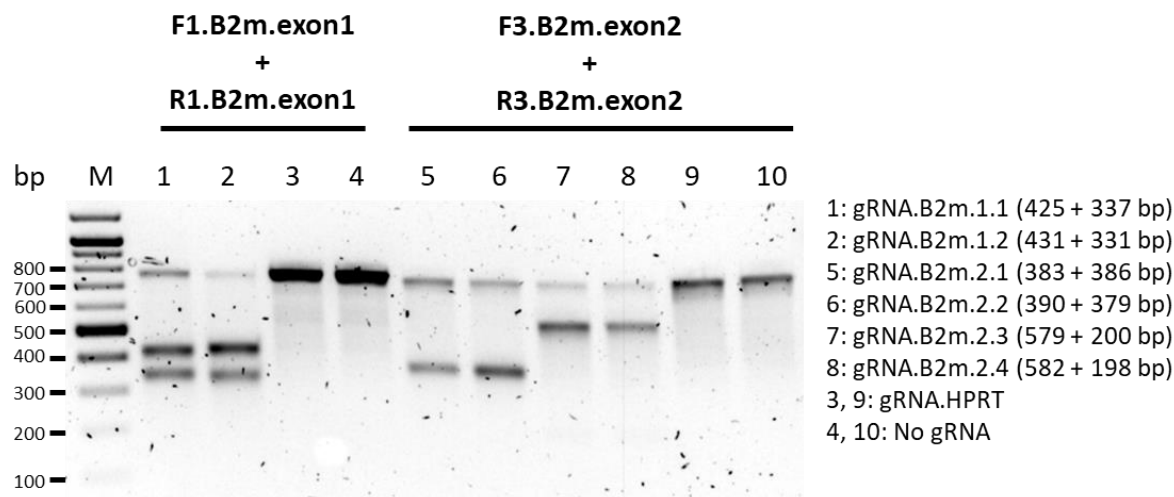


Figure 3.7. *In vivo* cleavage result of Cas9 RNPs targeting B2M.

Agarose electrophoresis results of T7EI digestion containing PCR amplicons from untreated HEK cells (lane 4 and 10) and HEK cells transfected with Cas9 RNP targeting *B2M* (lane 1, 2, 5-8) and *HPRT* (lane 3 and 9). The amplicons with indel mutations formed the heteroduplex and were digested by T7EI into smaller bands. The annotation (right) indicates the sgRNA used for each lane and corresponding expected cleavage products for each sgRNA. Data are presented of one experiment.

* *Small DNA bands diffused during electrophoresis*

To investigate how CRISPR-mediated knock-out of *B2M* affected cell surface expression of MHC-I, each HEK transfectant was analyzed by flow cytometry at day 1, 5 and 9 post-transfections. Cells were stained with the conformationally dependent W6/32 pan MHC-I specific antibody [333]. Flow cytometric analysis (Figure 3.8) revealed that W6/32-reactive MHC-I were expressed by untreated and *HPRT* targeted HEK cells from days 1-9. Flow cytometry indicated that surface MHC-I was detectable after day 1 by gene-edited cells. However, the numbers of cells expressing surface MHC-I complexes declined on days 5 and 9. Compared to other sgRNAs, sgRNA.B2M.2.2 led to the most significant decrease in MHC-I expression from 89.7% on day 1 to only 33.3% on day 5.

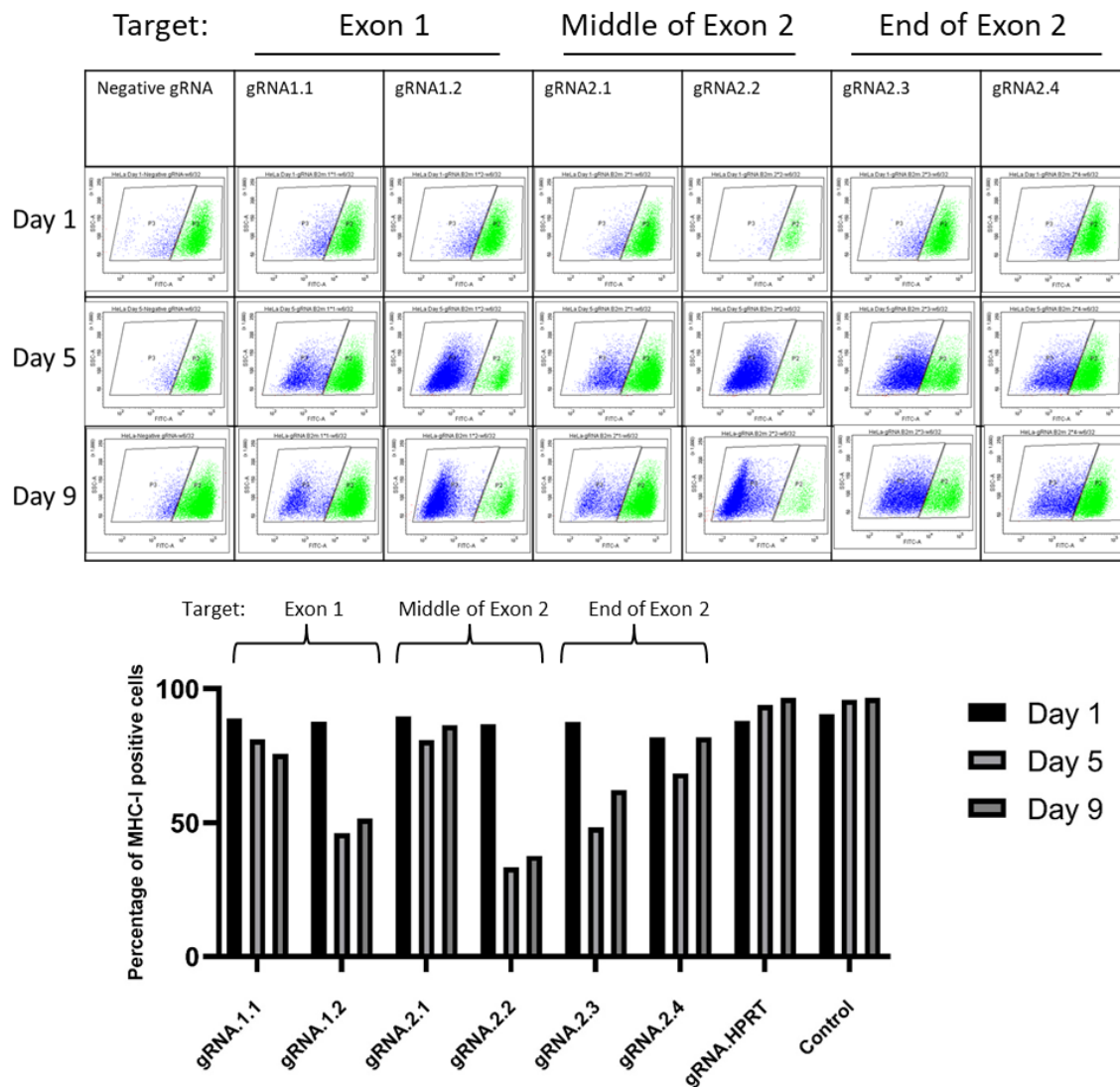


Figure 3.8. Investigation of surface MHC-I expression by gene-edited K562 cells.

Cell surface expression of MHC-I were analysed by FACS using the W6/32 antibody staining of HEK cells on 1, 5 and 9 day post-transfection with RNP targeting *B2M*. Untreated cells and cells treated with RNP targeting *HPRT* were used as a negative control. The FACS dot blots (top image) indicate the MHC-I positive (green) and MHC-I negative (blue) HEK populations. The bar graph (bottom image) indicate the percentages of MHC-I expression as determined by flow cytometry. Data are presented of one experiment.

HEK cells targeted with the RNP of sgRNA.B2M.2.2 were then used to isolate single cell clones which had a complete absence of MHC-I expression. Single-cell clones of gene-edited HEK cells were expanded and tested for MHC-I expression using W6/32 antibody staining. Flow cytometric analysis of single cell clones showed 3 in 11 clones had no expression of MHC-I: A5, B1 and C2 (Figure 3.9A). These clones were further expanded, and genomic DNA extracted for Sanger sequencing using the F3.B2M.exon1 primer. Sequencing results indicated that the HEK.B2M.KO.B1 cell line carried a deletion of 20 nucleotides on both B2M alleles, resulting in a frameshift mutation leading to the absence of B2M expression (Figure 3.9B). The HEK.B2M.KO.B1 cell line was therefore selected as the surrogate B2m deficient cell line for further experiments.

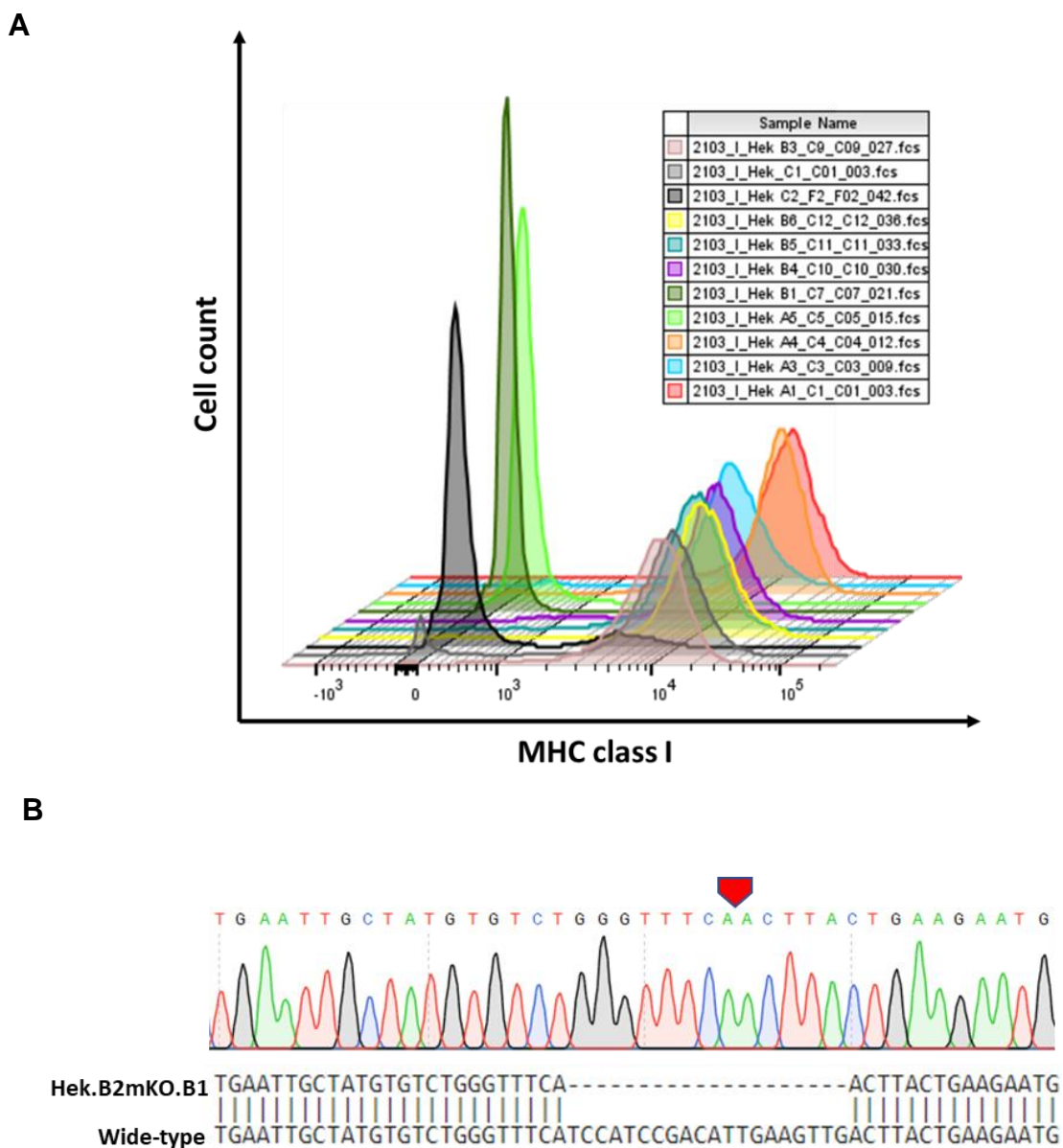


Figure 3.9. Identification of a homologous B2M knock-out HEK cell line.

(A) HEK cells transfected with RNP of sgRNA.B2M.2.2 were analyzed for cell surface expression of MHC-I molecules by flow cytometry using the W6/32 antibody. Flow cytometric data of each single clone were analyzed and merged using Flowjo software. (B) Sanger sequencing results confirmed the single clone B1 carried a homozygous deletion leading to the loss of B2M expression (NM_004048.4(B2M):del.152T→172G). Data are presented of one experiment.

3.3.5 Expression of PLB in B2M-deficient HEK cells

The HEKB2m deficient cell line was then employed to determine the function of the previously generated PLB by examining their ability to rescue surface expression of MHC-I. HEK.B2M.KO.B1 were transfected with the different pPLB plasmids listed in table 3.3. Wild type HEK cells and HEK.B2M.KO transfected with *pCR3.1-B2M* plasmid were used as a positive control for the expression of MHC-I. Untreated HEK.B2M.KO and HEK.B2M.KO transfected with an empty *pCR3.1* were used as a negative control for the expression of MHC-I. Transfected B2M-deficient cells were grown in selection medium (G418) followed by incubation with IFN- γ on day 5 to enhance expression of MHC-I. After d5, cell lysates were prepared and immunoblotted to assess B2M protein expression and flow cytometry analysis was performed to determine cell surface MHC-I expression.

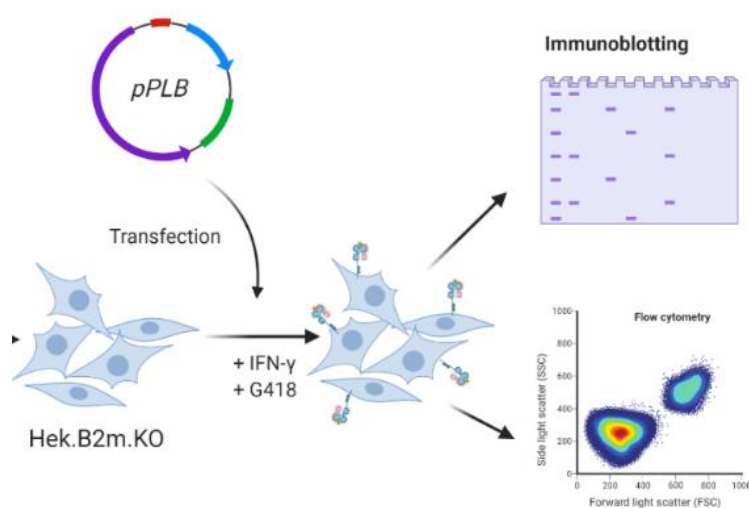
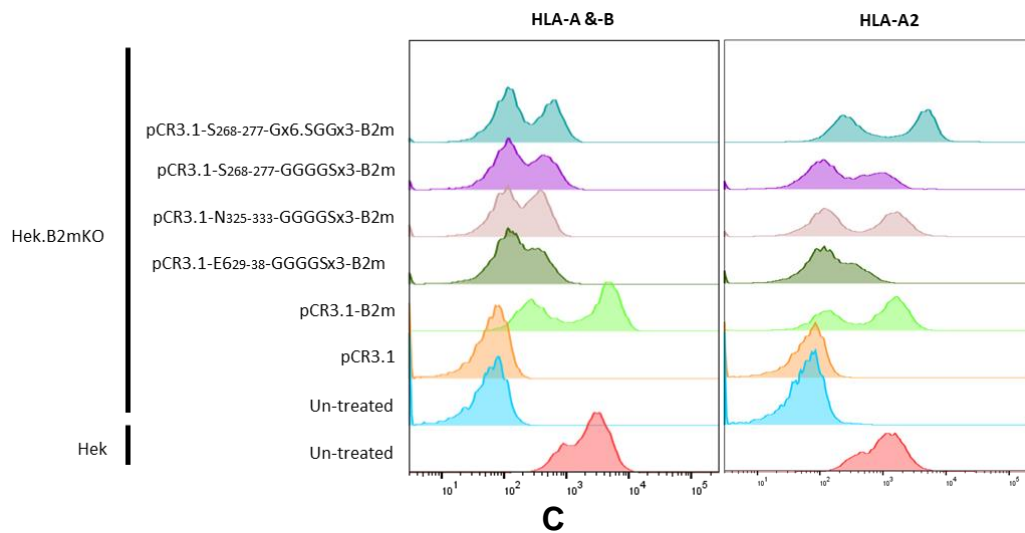


Figure 3.10. Strategy to evaluate function of pPLB plasmids.

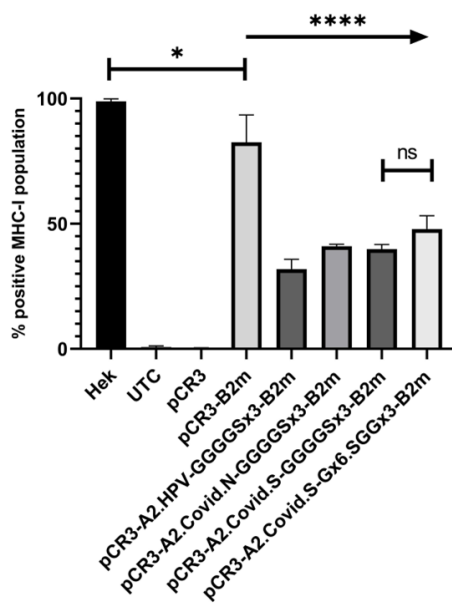
pPLB plasmids were transfected into B2M-deficient HEK cells, followed by growing in selection medium containing IFN- γ . The expressions of PLBs were evaluated by immunoblotting and flow-cytometry analysis on day 5 post-transfection.

To investigate how pPLB affected cell surface expression of MHC-I and HLA-A*02:01, each transfected cell was analyzed by flow cytometry using both W6/32 and the HLA-A2/-A28 specific BB7.2 antibodies [342]. The positive controls (HEK) demonstrated general MHC-I expression as determined by W6/32 staining and A2 expression as determined by BB7.2, while neither negative control (HEK.B2M.KO cells) exhibited any class I (Figure 3.11). W6/32 staining (Figure 3.11B) revealed that transfection of pPLB plasmids could rescue some expression of MHC-I on HEK.B2M.KO cells. However, The percentage of MHC-I positive population appeared on all *pCR3.1-peptide-linker-B2M* transfected samples is significantly lower compared to the positive control (HEK cells) as well as HEK.B2M.KO cells transfected with *pCR3.1-B2M* plasmid. BB7.2-stained flow cytometric analysis (Figure 3.11C) demonstrated that *pCR3.1-B2M* and different *pPLB* plasmids rescued HLA-A2 expression on HEK.B2m.KO to varying levels. Compared to *pCR3.1-B2M*, transfection of *pCR3.1-E629-38-(GGGGS)₃-B2M* , *pCR3.1-N325-333-(GGGGS)₃-B2M* or *pCR3.1-S269-277-(GGGGS)₃-B2M* plasmids could rescue HLA-A2 to a significantly lower percentage. Surprisingly, HEK.B2M.KO samples transfected with the *pCR3.1-S269-277-G6(SGG)₃-B2M* plasmid exhibited significantly higher rate of HLA-A2 positive population than positive control samples, suggesting that transfection of *pCR3.1-S269-277-G6(SGG)₃-B2M* significantly enhanced HLA-A2 expression.

A



B



C

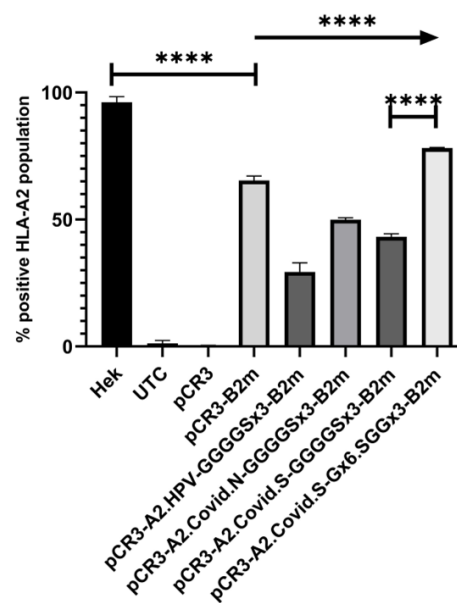


Figure 3.11. Flow-cytometry analysis of B2M-deficient cell transfected with pPLB plasmids.

(A) The flow cytometry histograms of surface MHC-I (A, B & C) and HLA-A*02 expression by wild type, untreated B2M-deficient and B2M-deficient HEK cells transfected with pCR3.1 and pPLB plasmids. Surface MHC-I and HLA-A*02 expression was analyzed by flow cytometry using W6/32 (left) and BB7.2 antibodies (right), respectively. (B) Bar charts indicating the percentage of positive MHC-I population as determined by W6/32 staining. (C) Bar charts indicating the percentage of positive HLA-A*02 expression. Two-

way ANOVA. **** $P < 0.001$. The means of MFI were calculated from triplicate data points (mean \pm SEM, $n=3$).

To examine the expression of PLB proteins by the various B2M-deficient HEK transfectants, equal number (10^6) of cells were lysed in 1% NP-40 detergent. Reduced lysates were resolved on 17% SDS-PAGE followed by immunoblotting with anti-B2M antibody to assess B2M protein expression, and anti-GAPDH antibody as an internal loading control. Immunoblotting with GAPDH (Figure 3.12, top row) revealed approximately equal protein loading between the cell lysates. Immunoblotting with anti-B2M antibody (Figure 3.12, bottom row) demonstrated that the wild-type (wt) B2M proteins (11.7 kDa) were observed in cell lysate from WT HEK cells (lane 1) but not in either untreated B2M-deficient HEK cells (lane 2) or B2M-deficient HEK transfected with the empty pCR3.1 vector (lane 3). B2M was detected in cell lysates of B2M-deficient HEK transfected with the pCR3.1-B2M plasmid (lane 5). The calculated protein sizes of PLB proteins are slightly different depending on the inserted peptide and linker sequence. Using ExPASy to calculate predicted protein molecular weights, PLB-N₃₂₅₋₃₃₃-(GGGGS)₃, PLB-S₂₆₉₋₂₇₇-(GGGGS)₃ and PLB-S₂₆₉₋₂₇₇-G6(SGG)₃ were expected to have a Mw of approximate 13.6 kDa, whilst that of PLB-E₆₂₉₋₃₈-(GGGGS)₃ was expected to be larger with a size of 13.8 kDa. Although bands with the corresponding sizes to the different PLB proteins were detected by immunoblotting with anti-B2M antibody, the intensity of these bands were different. Compared to the wt B2M band, PLB-S₂₆₉₋₂₇₇-(GGGGS)₃ and PLB-S₂₆₉₋₂₇₇-G6(SGG)₃ demonstrated lower expression levels but PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ appeared to exhibit higher expression levels. These observations suggests that the type of peptides and linker influence the cellular capacity to synthesize PLB molecules.

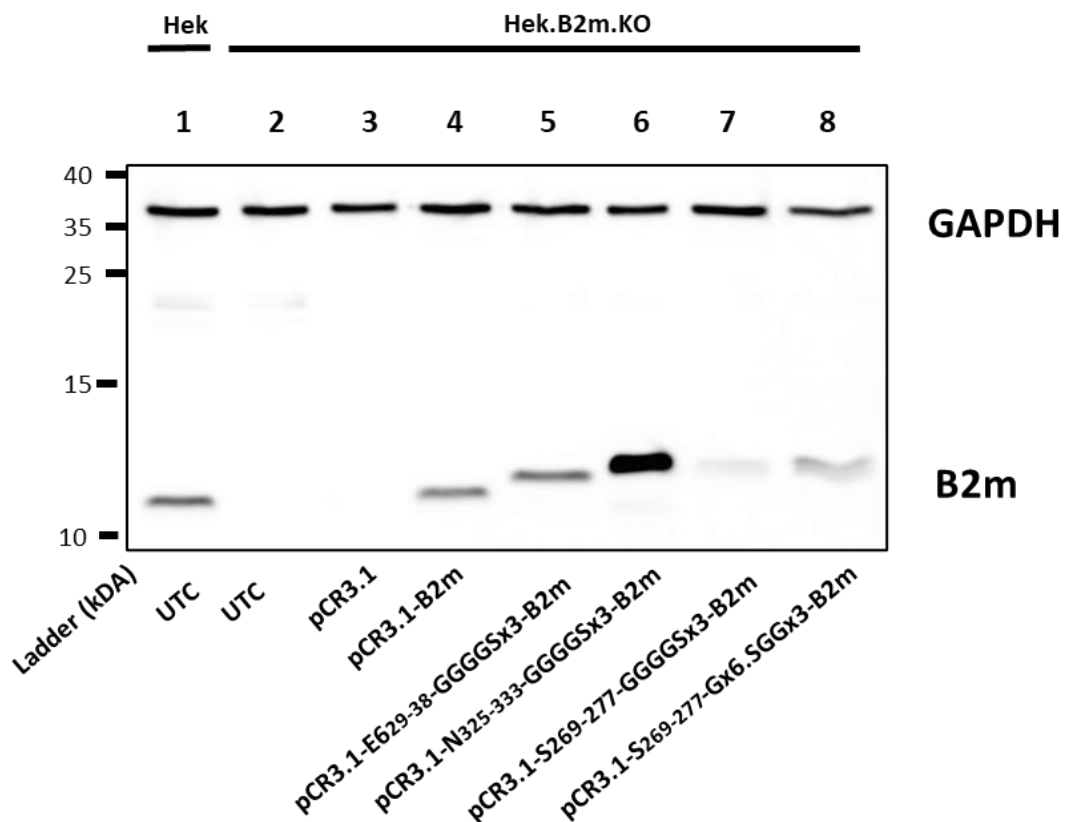


Figure 3.12. Immunoblotting of B2M and PLB proteins.

Wild type HEK and B2M-deficient cells transfected with pCR3.1 and pPLB plasmids were lysed in 1% NP-40 detergent. Reduced lysates were resolved by 17% SDS-PAGE gel followed by immunoblotting with anti-GAPDH and anti-B2M antibodies. Cell lysates of HEK were loaded in lane 1 and those of HEK.B2M.Ko were loaded in lane 2-8. Lane 2 is untreated B2M-deficient HEK cells. Lanes 3, 4 and 5-8 are lysate samples of B2M-deficient HEK cells transfected with pCR3.1, pCR3.1-B2M and different pPLB plasmids, respectively.

3.3 Discussion

In this part of the thesis, I describe the first PLB molecules containing peptides derived from the SARS-CoV-2 virus, which is the pathogen leading to Coronavirus disease 2019 (COVID-19). Toward this goal, I have created three SARS-CoV-2 PLB and one HPV PLB that were transfected into B2M-deficient cells to evaluate their ability to rescue MHC-I expression prior to use in gene editing experiments.

There have been previous studies with similar Peptide-linked B2M (PLB) molecules which have been referred to as epitope-linked B2-microglobulin [343] or peptide-B2M fusion molecules [344]. PLBs could generate peptide-specific CTL responses and potentially become an immunotherapy or vaccine candidate to combat many neoplastic diseases or viral infections. The only B2M-deficient cell line used for testing PLB was the human colorectal adenocarcinoma DLD-1 cells, which expresses two mutations on two B2M alleles (each allele carrying a separate mutation) [345], and reported in Tafuro's study [344]. Tafuro's group transduced the retroviral vector of HLA-A2-PLB containing the influenza matrix epitope (GILGFVFTL) into DLD-1 cells to generate cells expressing an MHC complex of HLA-A2 and the fusion protein. The result of ⁵¹chromium-release assays showed that flu-matrix -specific CTLs can lyse 78% of DLD-1 cells expressing the PLB.

In my study, 6 different sgRNAs for CRISPR/Cas9 were used for knocking-out B2M expression within the HEK cell line which also expresses HLA-A2. The *IVT* and *IVV* cleavage assays, demonstrated that all RNP of Cas9 nuclease and sgRNA have the ability to generate DSBs at the selected target sites. On day 5 post-transfection, the negative MHC-I HEK population could be separated clearly from the positive MHC-I population (Figure 3.11A). The monoclonal B2M-deficient HEK cells were isolated

from the polyclonal population and confirmed by Sanger sequencing as well as immunoblotting with anti-B2M antibody. The CRISPR/Cas9-mediated system has brought a huge advantage to precisely knock-out human *B2M*.

The expressions of MHC-I on B2M-deficient HEK cells were rescued after transfection with different pPLB plasmids. These constructs have been shown to rescue MHC-I but at a much lower expression level compared to wild type HEK cells or positive control (B2M-deficient HEK transfect with PCR3.1-B2M vector). On the other hand, the expression of HLA-A2 was detected at just slightly lower or similar levels between transfected cells and control. Especially, in the case of the HEK.B2M.KO cell line transfected with the pCR3.1-S₂₆₉₋₂₇₇-G6(SGG)₃- *B2M* plasmid, expression of HLA-A2 molecules at the cell surface was enhanced compared to *pCR3.1-B2M* vector. This observation suggests that cells can synthesis the PLB protein from the PLB plasmid and the PLB protein can form the MHC complex with only its corresponding peptide-specific HLA molecule.

Combined with the B2M immunoblotting data, it suggests all PLB molecules have been successfully translated and transported into the ER, possibly only binding to the specific HLA class I heavy chain and then translocated to cell membranes in the form of MHC-I – PLB complex. For future analysis, the PLB transported to the plasma membrane could be detected using immunoblotting of PLB protein released by cells in the growth medium or by cell surface biotinylation experiments. Other HLA class I heavy chains might still bind to the PLB but with the A2-specific peptides not associating with the binding groove. Additionally, the PLB molecules could bind to non-HLA-A2 heavy chains and prevent other endogenous peptides from binding. Without

peptides, MHC I molecules do not pass the ER quality control system and cannot be transported to the cell surface [73].

Protein linkers play an important role in the development of novel biomolecules with different functional properties. Directed-fusion proteins without a linker or with an inappropriate linker may lead to unexpected outcomes, including impaired bioactivity [240], low yield in protein production [241], or misfolding of the recombinant protein [242]. Based on their functionalities, linkers can be classified into three types: flexible, rigid and cleavable. While flexible linkers have the advantage of joining functional domains which require a certain degree of interaction or movement, rigid linkers are frequently used to maintain the relative stiff structure between domains. Cleavable linkers are not stable and allow for *in vivo* separation of domains. In both the SCT and PLB, flexible linkers are the most appropriate to span the gap between the N terminus of B2M and the C terminus of the peptide. Based on previous reports, two linkers, (GGGGS)₃ and G₆(SGG)₃, were used to develop pPLB plasmids containing SARS-COV-2 S₂₆₉₋₂₇₇ peptide and B2M [346, 347]. Interestingly, although both linkers have the same numbers of Gly and Ser residues, they provided distinct contributions to the expression of peptide-specific HLA class I at the cell surface. Indeed, the level of expression of HLA-A2 on HEK.B2M.KO transfected with pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M plasmid was 5 times higher than that of HEK.B2M.KO transfected with pCR3.1-S₂₆₉₋₂₇₇-(GGGGS)₃-B2M plasmid. Different PLB constructs with different peptides need to be tested to determine how (GGGGS)₃ and G₆(SGG)₃ linkers really impact on the expression of PLB at the cell surface. Additionally, the length of (GGGGS)_n and G₆(SGG)_n linkers can be changed by adjusting the copy number “n” to maintain

necessary interaction between peptide and binding groove pockets of MHC-I complexes.

Since the first description in 2009 by Dr. Daniel Gibson and colleagues [348], Gibson Assembly (GA) has been widely applied for the assembly of multiple linear DNA fragments in a single isothermal reaction. In this study, linear pCR3.1- *B2M* plasmid and DNA fragments containing desired peptides and linker were fused together using GA cloning master mix consisting of three enzymes: T5 Exonuclease, DNA polymerase and Taq DNA Ligase. After achieving the first pPLB plasmids, I developed a short GA cloning protocol to construct the new PLB plasmid which required the same linker but alternative peptides. In this short GA protocol, the pPLB template was linearized using a pair of 60nt primers containing the new peptide sequences. The short GA protocol only required a pair of primers to accomplish construction of the pPLB plasmid instead of the gBlock or Megamer DNA fragment required for conventional GA. As a result, the short GA protocol can be cost saving for developing future pPLB plasmids as well as introducing different peptides into SCT constructs.

In summary, our result demonstrated that the SARS-CoV-2 peptide-B2M fusion molecules can be expressed and rescue the cell surface expression of HLA-A2. While (GGGGS)₃ are routinely used to develop both PLB and SCT, the expression of MHC-I shows that G₆(SGG)₃ linker can enhance the amount of HLA-A2 at the cell surface. However, this analysis was performed with a limited number of peptides. It would be of interest to determine if these molecules could induce CTL killing and whether these molecules proceeded through the usual quality control mechanisms governing MHC-I expression.

CHAPTER 4
**GENERATING PEPTIDE-LINKED B2M USING
CRISPR/CAS9-MEDIATED KNOCK-IN**

Chapter 4 Generating peptide-linked B2M using CRISPR/Cas9-mediated knock-in

4.1 Chapter aims & objectives

The development of the peptide-linked B2M (PLB) and the MHC I single-chain trimer (SCT) have been considered as potentially important DNA vaccines against tumors [182, 187, 344] and pathogens [189, 349]. Up-to-now, most SCT and PLB constructs were delivered into cells via viral or non-viral platforms. Viral vectors are a promising system of gene delivery because of their efficiency and specificity which can target either a wide range of cells or specific cell types. However, the limitation of viral vectors includes risks of replication-competent virus, tumorigenesis, and pre-existing immunity. On the other hand, non-viral vectors or plasmids are cost-saving and can exhibit low immunogenicity, but delivery such as transfection of plasmids can be highly toxic and expression of desired complexes can be unstable. In this study, I propose that integration of peptide and linker into the *B2M* locus can lead to and maintain the stable expression of PLB molecules by cells. Due to the low immunogenicity and limited off-target effect, CRISPR/Cas9 Ribonucleoprotein (RNP) and donor templates will be used to introduce the DSBs at the target exon of B2M to facilitate HDR of peptide and linker for the generation of cell lines stably expressing PLBs.

To achieve the above aims, several objectives were established, including:

1. Designing and optimizing CRISPR/Cas9 systems and donor templates to introduce the PLBs into the endogenous B2M gene.
2. Generation of cell models stably expressing PLB molecules at the cell surface and characterization of PLB expression using flow cytometry and immunocytochemistry.
3. Determination of PLB function by CD8+ effector T cells

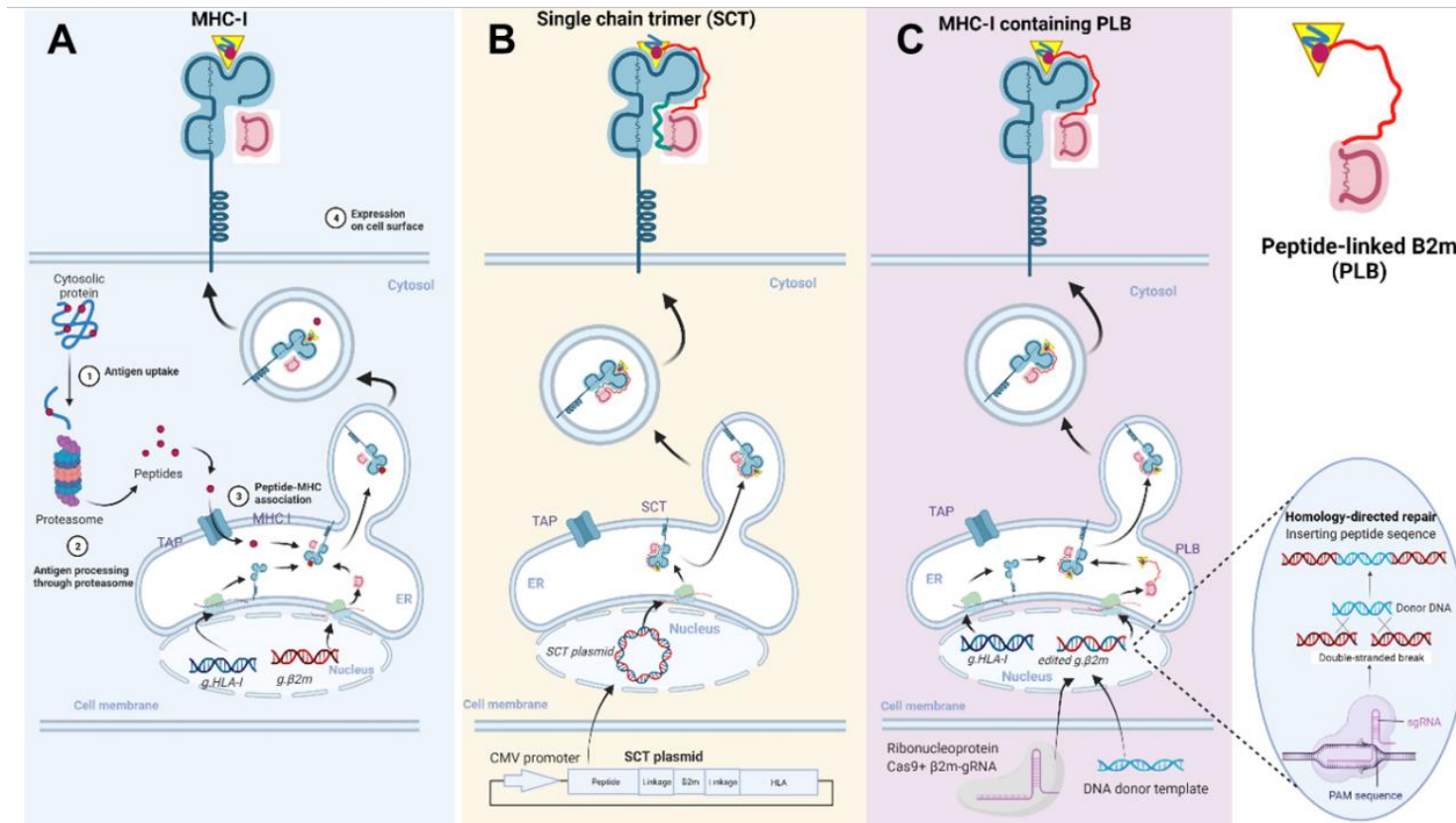


Figure 4.1. Antigen presentation pathways of MHC-I, SCT and MHC-I-PLB complex.

(A) Cytoplasmic proteins are degraded by proteasomes to peptides. The peptides are transported into the endoplasmic reticulum and assembled with class I MHC molecules. The peptide-class I MHC complexes are transported to the cell surface and are ready to be screened by CD8⁺ T cells. (B) SCT construct containing all MHC-I components fused by covalent linkages can be expressed on the plasma membrane and present specific peptides to T cells. (C) Using CRISPR/Cas9, specific peptide and linkage sequences are directly inserted into endogenous *B2M* to encode PLB molecules which can stably present the antigen on MHC class I molecules.

4.2 Results

4.2.1 CRISPR/Cas9 system targeting exon 1 of B2M

The first aim of this objective was to design and optimize a CRISPR/Cas9 system which could generate DSBs at the end of the *B2M* exon 1. To facilitate HDR of PLB within the genomic B2M sequence, the original allele had to be cleaved by Cas9 RNP to generate single or double stranded breaks and be repaired using a donor DNA template. By using the bioinformatics tool- Benchling, single guide RNAs were designed to target the end of exon 1 of B2M (Figure 4.2). The gRNA.B2M.1.3, had an AGG PAM sequence to enable the generation of a DSB at the location of -11 nt from the last nucleotide of exon 1.

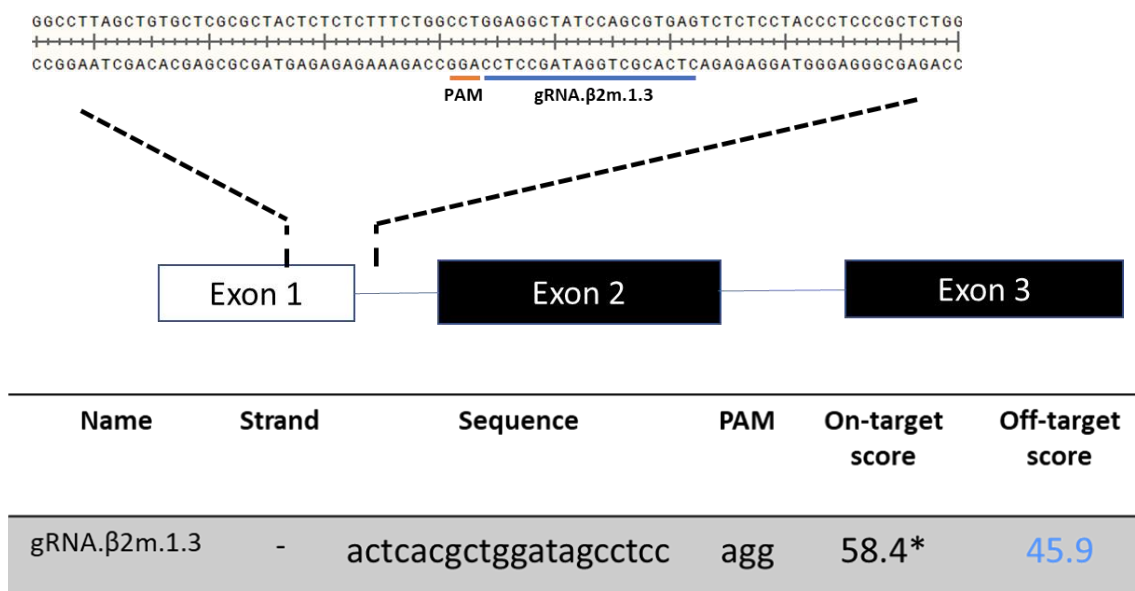


Figure 4.2. Design of sgRNA targeting the end of B2M.exon_1.

Genomic *B2M* contains 3 exons located on the long arm of chromosome 15 at 15q21.1. sgRNA.B2M.1.3 (blue) was designed to target the end of exon 1 which contains the intersection of the signal peptide and sequences of the B2M protein. 20-nt crRNAs follow the PAM motif (red) at their 3' end by a 5'- NGG motif. The table (bottom) stated the detail of the designed sgRNA.

IVT and IVV cleavage assays were performed to validate the suitability of sgRNA.B2M.1.3. Previously sgRNA.B2M.1.2 was shown to generate DSB on exon 1 of B2M (see Figure 3.6 and 3.7) and was therefore used as a positive control. Human *HPRT* positive sgRNA (IDT) were used as a negative control for cleavage of the B2M amplicon.

The naked PCR product of B2M.exon1 amplified by F1.B2M.exon1 and R1.B2M.exon1 was treated with Cas9 RNP of sgRNAs followed by electrophoresis separation on a 1.2% agarose gel. *IVV* cleavage results demonstrated that Cas9 RNP of sgRNA.B2M.1.3 could cleave naked PCR amplicons (762 bp) of B2M.exon 1 generating a 458 bp and a lower Mw band of 304 bp (lane 2, Figure 4.3A). Subsequently, Cas9 RNPs were transfected into HEK cells by using Lipofectamine CRISPRMAX Cas9 Transfection Reagent (ThermoFisher). After 24 hours, transfected cells were harvested for gDNA extraction followed by PCR amplification of B2M.exon1. The PCR amplicons were denatured and reannealed to allow the formation of heteroduplexes between mutated DNA and wt DNA followed by the treatment with T7 endonuclease I. The cleaved products were separated by gel electrophoresis as described. Electrophoresis image of the *in vivo* cleavage assay (Figure 4.3B) showed that cleaved bands (458 and 304 bp) were observed by samples treated with Cas9 RNP of sgRNA.B2M.1.3 (lane 2). These results indicate sgRNA.B2M.1.3 can guide the Cas9 protein to the target site where the RNP induces the DSB on genomic B2M and facilitates the cell repair process following NHEJ mechanisms.

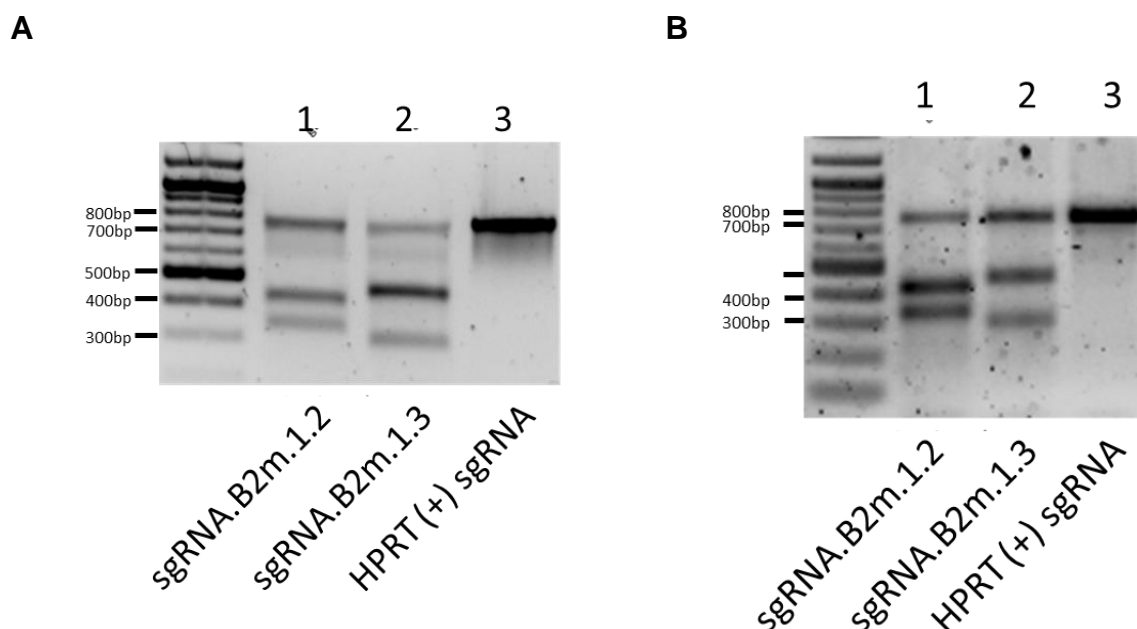


Figure 4.3. Functional assays of sgRNA.B2M.1.3.

Electrophoresis results indicate the: (A) *In vitro* cleavage assay and (B) *In vivo* cleavage assay of RNP of Cas9 protein and sgRNA.B2M.1.3 (lane 2). Samples transfected with RNP of sgRNA.B2M.1.2 (lane 1 and HPRT positive sgRNA were used as positive and negative controls, respectively). Amplicons of *B2M* exon 1 have the length of 769 bp. Cleaved bands of RNP.B2M.1.2 and RNP.B2M.1.3 were expected to generate products of 431+331 bp and 458+304 bp, respectively. Data are presented of one experiment.

The *IVT* and *IVV* cleavage assays demonstrated that the designed sgRNA.B2M.1.3 should be successful in guiding Cas9 protein to the *B2M* exon 2 and facilitate the DSB at this site. In order to check the indel mutations caused by Cas 9 RNP of sgRNA.B2M.1.3 in transfected HEK cells, the short amplicon of the target region was generated by F- OT.sgB2M1.3_1 and R- OT.sgB2M1.3_1 (Appendix 5) and examined by deep sequencing using Illumina NGS. By comparison with the reference genomic *B2M* sequence, NGS data analysis (Figure 4.4) revealed that there were a wide range of different indel mutations introduced into the *B2M* locus of gene-edited HEK cells.

The most outstanding dominant mutations (>2%) were a frame shift 8bp deletion (5.68%), frame shift 1 bp deletion (4.85%), in frame silent variant of a 7 bp substitution (3.07%), and frame shift 1 bp insertion (2.54%).

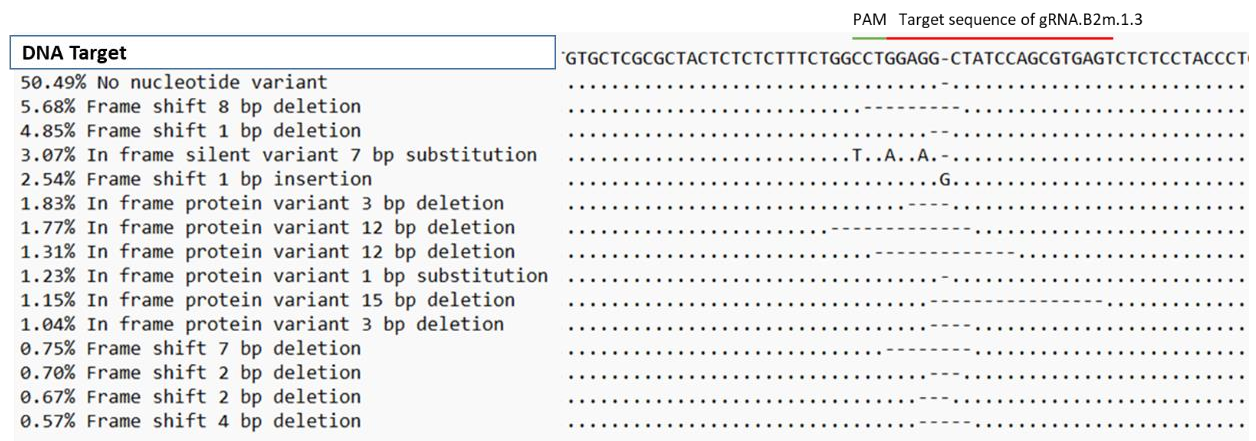


Figure 4.4. INDEL spectrum of sgRNA.B2M.1.3 targeting the *B2M* locus.

Sequencing data indicates the INDEL spectrum of sgRNA.B2M.1.3 in the targeted B2M locus. The top row shows the reference sequence of the targeted B2M region, which contains crRNA (red) and its PAM (green). The bottom rows show a broad range of INDELS on the aligned deep-sequencing data of HEK transfected with Cas9 RNP of sgRNA.B2M.1.3. Sequencing data were aligned with reference sequences using Geneious software.

4.2.2 Optimization of HDR to insert peptide and linkers into the endogenous *B2M* locus

After achieving optimized sgRNAs for introducing DSBs within the exon 1 *B2M* locus, donor templates containing the desired peptides and linkers were designed to knock-in these sequences into the target site located at the 3' end of the first 60 nucleotides of *B2M*. Double-stranded DNA (dsDNA) was used as a donor template for HDR on gene-edited cells. Three different peptide sequences were used, which included, HPV E6₂₉₋₃₈, SARS-CoV-2 N₃₂₅₋₃₃₃, and SARS-CoV-2 S₂₆₉₋₂₇₇. These peptides were designed with two different linker sequences i.e. (GGGGS)₃ and G₆(SGG)₃. The DNA donor template sequence of peptide-linkers, must also contain two homology arms flanking the target site [203]. In order to knock-in a 72-75 nt sequence insertion, I designed two types of homology arm which varied in length, one containing 60 nt (dsODN) and another of 200 nt (referred to as a Megamer) (Figure 4.5). Full sequences of all donor templates are listed in Supplementary Data 3 (Appendix 3).

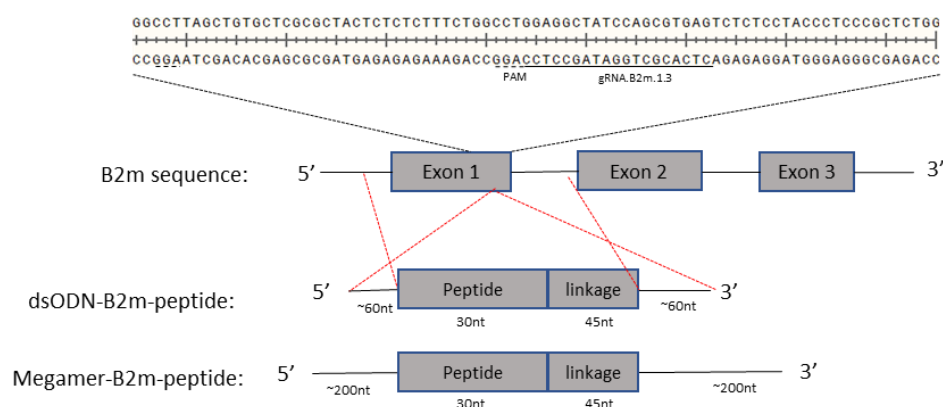


Figure 4.5. Design of donor templates for HDR of PLB.

Donor templates can be used in combination with Cas9 RNPs to introduce HDR-mediated insertion at the target site of exon 1 *B2M* locus. The insertion sequence included peptide

and linker with a length of 72-75 nt. The homology arm of dsODN and Megamer donor templates have the length of 60 and 200 nt, respectively.

At the early stage of this project, I used Cas9 RNP of sgRNA.B2M.1.3 and dsODN.E6₂₉₋₃₈-(GGGGS)₃ to test the validity of PLB insertion. The HDR on the B2M locus was facilitated by co-transfected dsODN donor templates and RNP of Cas9 protein and sgRNA.B2M.1.3 using electroporation with the Amaxa Nucleofector system (Lonza). Equal numbers (10^6) of HEK and K562 cells were transfected with ssODN and RNP complexes at the molar ratio of 4:1 followed by the incubation in complete culture media with HDR Enhancer. Genomic DNAs were extracted from day one post-transfected cells and used for preparing amplicons for the T7EI assay by F1.B2M.exon1 and R1.B2M.exon1 primers. Following resolution by DNA gel electrophoresis, the T7EI assay (bottom image, Figure 4.6) demonstrated the cleaved bands of 458 and 304 bp can be detected in both K562 and HEK derived samples transfected with Cas9/sgRNA.B2M.1.3 RNP (Figure 4.6, lane 1, 2, 4 and 5). This result indicates that Cas9/sgRNA.B2M.1.3 RNP complexes can introduce DSBs in all transfected cells and can generate both indel mutations (without donor template) and the desired insertion (with donor template). Electrophoresis result of PCR products (top image, Figure 4.6) show a band of 837 bp arising from HDR (red arrow) in K562 cells transfected with RNP and donor template. This data indicates the dsODN template can introduce the insertion of PLB in K562 but not in HEK cells

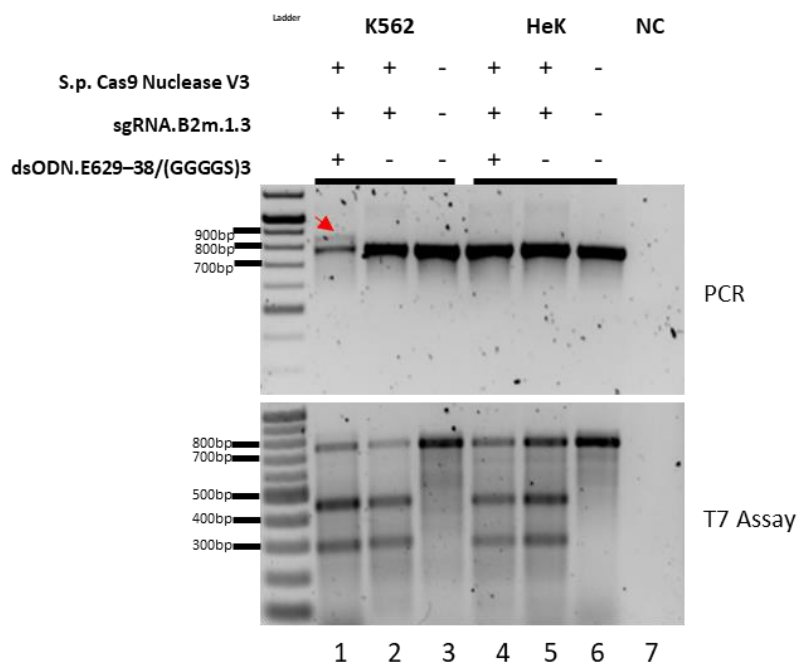


Figure 4.6. Detection of HDR of PLB insertion in K562 and HEK cells

DsODN donor templates of HPV E6₂₉₋₃₈ peptide and (GGGGS)₃ linker and RNP.B2M-1.3 were co-transfected into K562 and HEK cells. At one day post-transfection, PCR amplicons of transfected cell-derived genomic DNA were used to check HDR efficiency by PCR (top gel) and T7EI assay (bottom gel). The size of wt B2M amplicon and HDR amplicon were 762 and 837 bp. Data are presented of one experiment.

Although the HDR of the PLB using dsODN as a donor template were detected in K562 cells, this cell line does not express the HLA-A*02 allele which is required for the expression of the designed PLB constructs. Therefore, the HDR experiment was repeated as above using the Megamer donor template (Megamer.E6₂₉₋₃₈-(GGGGS)₃) instead of dsODN to overcome the possible HDR obstacle within the HEK cell line. Following PCR and T7EI assay (Figure 4.7), the HDR band of 837 bp (lane 1, red arrow) and T7EI-cleaved bands of 458 and 304 bp (lane 3) were detected in HEK cells transfected with RNP.B2M.1.3 and Megamer.E6₂₉₋₃₈-(GGGGS)₃. These results

indicated that HDR of the PLB was successful in HEK cells transfected with Cas9 RNP and Megamer.

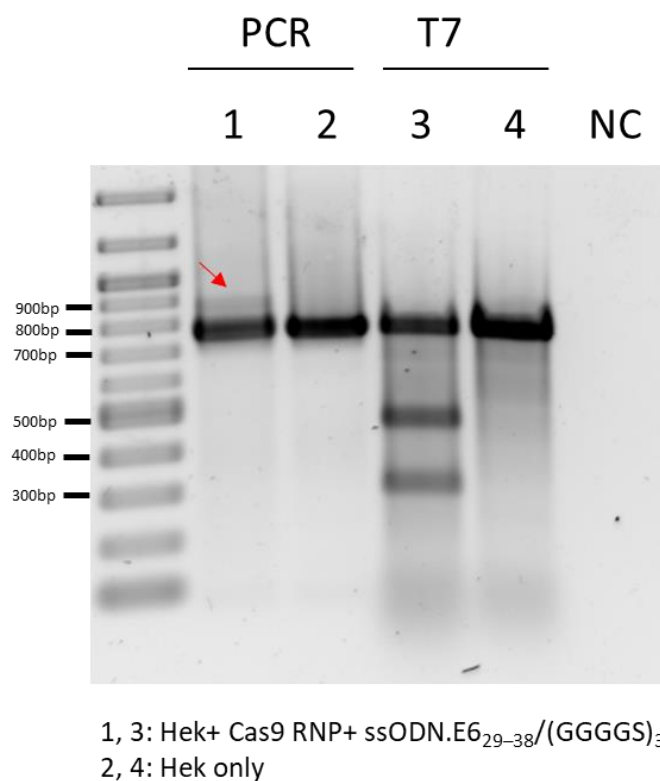


Figure 4.7. Evaluating insertion of peptide and linker into the endogenous B2M allele on HEK cells.

Megamer donor templates of HPV E6₂₉₋₃₈ peptide and (GGGGS)₃ linker and RNP.B2M-1.3 were co-transfected into HEK cells. At one day post-transfection, PCR amplicons of transfected cell-derived genomic DNA were used to check HDR efficiency by the T7E1 assay. The size of wt B2M amplicon and HDR amplicon were 762 and 837 bp. Data are presented of one experiment.

In summary, I have successfully introduced the peptide and linker sequences into exon 1 of the *B2M* locus by using CRISPR/Cas9 and dsODN/Megamer donor templates. The gene-edited HEK cells were expected to produce a PLB protein which carries the

specific peptide for binding to the endogenous HLA-A*02 allele. This method was later used to generate different gene-edited cells for evaluating the expression of PLBs.

4.2.3 Expression of PLB-E6₂₉₋₃₈-(GGGGS)₃ proteins in HEK cell

Following transfection with the donor template E6₂₉₋₃₈/(GGGGS)₃ and Cas9 RNP of sgRNA.B2M.1.3, gene-edited HEK cells were in a heterogenous cell population (Figure 4.7). Therefore, monoclonal cell lines were generated from polyclonal pools of stably transfected cells using single-cell isolation (see materials and method section 2.1.4). After 7-10 days, single-cell clones of gene-edited HEK cells were expanded and partly harvested for extraction of genomic DNA. 41 clones of HEK cells were isolated and expanded. PCR amplicons of each single-cell clone using F1.B2M.exon1 and R1.B2M.exon1 primers were used to identify monoclonal cells containing desired mutations based on the size of amplicons that would be generated from wt B2M (762 bp) and HDR B2M modified sequences (837 bp). PCR analysis demonstrated that only one in 41 HEK clones (2.4%) potentially contained the desired insertion of E6₂₉₋₃₈/(GGGGS)₃ (lane 7, Figure 4.8A). Sanger sequencing was performed using DNA extracted from clone 7 (Figure 4.8B) which indicated that exon 1 of the *B2M* locus of the monoclonal HEK-E6₂₉₋₃₈/(GGGGS)₃ cell line, contained a 100% match with the sequence of the Megamer-E6₂₉₋₃₈-(GGGGS)₃ donor template (Figure 4.8 B).

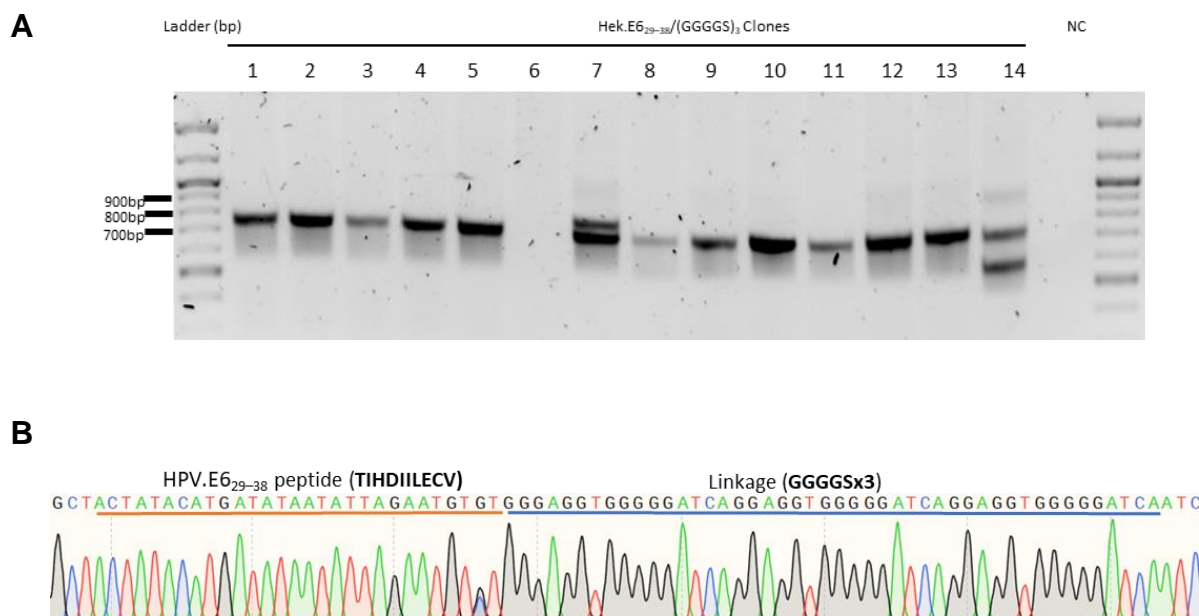


Figure 4.8. Single-cell isolation of monoclonal HEK cells carrying E6₂₉₋₃₈/((GGGGS)₃.

Monoclonal gene-edited cells were isolated using serial dilution of polyclonal populations followed by DNA extraction for PCR with F1.B2M.exon1 and R1.B2M.exon1 primers. (A) DNA electrophoresis image of PCR product using genomic DNA templates extracted from single cell HEK clones. The lengths of wt and gene-edited PCR carrying E6₂₉₋₃₈/((GGGGS)₃ are 762 and 837 bp, respectively. (B) Sanger sequencing results confirmed the desired insertion of E6₂₉₋₃₈/((GGGGS)₃ into exon 1 of B2M allele. Data are presented of one experiment.

To investigate how CRISPR-mediated insertion of peptide and linker within the B2M locus affected cell surface expression of MHC-I molecules, the monoclonal gene-edited HEK-E6₂₉₋₃₈/((GGGGS)₃ cells were analyzed by flow cytometry using W6/32 and BB7.2 antibodies. Prior to antibody staining, un-transfected and gene-edited clones were incubated with IFN- γ (400 IU/mL) for 3 days to enhance expression of MHC-I. Experiments were performed in triplicate and the average MFI calculated (\pm Standard Error of Mean (SEM)). Flow-cytometry data of each sample stained by the same antibody were compared using unpair t-test (Prism) (Appendix 4, supplement data 6

and 7). Flow cytometric analysis (Figure 4.9) revealed that W6/32 and BB7.2- reactive HLA types were expressed in both HEK and HEK-E6₂₉₋₃₈/(GGGGS)₃. However, the average MFI of W6/32 and BB7.2 staining (14090 ± 89 ; 7010 ± 163 , respectively) of HEK-E6₂₉₋₃₈/(GGGGS)₃ were significantly lower than those of WT HEK cells, suggesting gene-edited HEK-E6₂₉₋₃₈/(GGGGS)₃ had very limited expression of HLA-A2 on the cell surface.

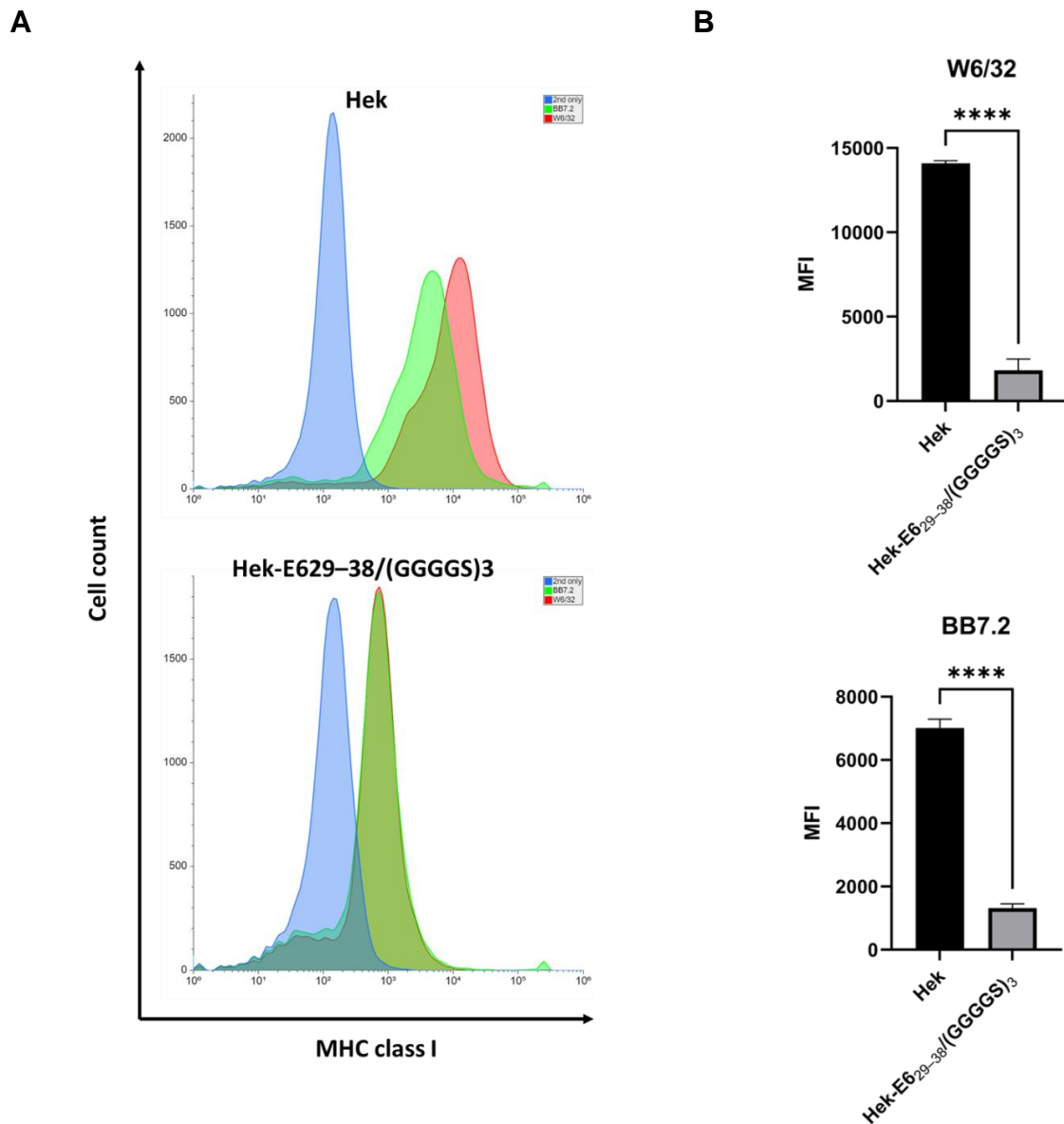


Figure 4.9. Flow-cytometry analysis of MHC-I expression by HEK-E6₂₉₋₃₈/(GGGGS)₃ cells.

Surface MHC-I and HLA-A*02 expression was analyzed by flow cytometry using W6/32 and BB7.2 antibodies, respectively. (A) The flow cytometry histograms of surface MHC-I and HLA-A*02 expression by HEK (top) and HEK-E6₂₉₋₃₈/(GGGGS)₃ (bottom). (B) The bar charts indicate the Mean Fluorescence Intensities (MFI) of MHC-I (top) and HLA-A*02 (bottom) expression by HEK and HEK-E6₂₉₋₃₈/(GGGGS)₃. One-way ANOVA with Turkey's post-test with ****P<0.001. The means of MFI were calculated from triplicated data (mean ± SEM, n=3).

To examine the expression of PLB- E6₂₉₋₃₈/(GGGGS)₃ protein by transfected B2M-deficient HEK cells, equal number (10⁶) of cells were lysed in 1% NP-40 detergent. Reduced lysates were resolved on 17% SDS-PAGE followed by immunoblotting with anti-B2M antibody and anti-Protein disulfide-isomerase (PDI) antibody, which probed for PDI protein as an internal loading control. Immunoblotting with PDI (Figure 4.10, bottom row) demonstrated approximate equal amounts of protein between the cell lysates. Immunoblotting with anti-B2M antibody (Figure 4.10, top row) demonstrated wild-type (wt) B2M proteins (11.7 kDa) in cell lysates from WT HEK cells (lane 5) and HEK-E6₂₉₋₃₈/(GGGGS)₃. The PLB-E6₂₉₋₃₈/(GGGGS)₃ protein, with an expected size of 13.8 kDa was not detected in cell lysates derived from HEK-E6₂₉₋₃₈/(GGGGS)₃ transfected cells. These observations suggests that the PLB-E6₂₉₋₃₈/(GGGGS)₃ protein was not successfully synthesized by the gene-edited cells.



Figure 4.10. Detection of B2M and PLB by HEK-E6₂₉₋₃₈-(GGGGS)₃ cells.

Wild type HEK and HEK-E6₂₉₋₃₈-(GGGGS)₃ cells were lysed in 1% NP-40 detergent. Reduced lysates were resolved by 17% SDS-PAGE gel followed by immunoblotting with anti-PDI and anti-B2M antibodies. Cell lysates of wt HEK were loaded in lane 1 and those of HEK-E6₂₉₋₃₈-(GGGGS)₃ cells were loaded in lane 2. WT B2M (11.7kDa) was detected whilst PLB-E6₂₉₋₃₈/(GGGGS)₃ protein (13.8 kDa) was not apparent. Expression of PDI (57kDa) demonstrated approximate equal loading. Data are presented of one experiment.

4.2.4 Expression of PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ on HEK

Using the HPV E6₂₉₋₃₈ peptide, the PLB construct with a (GGGGS)₃ linker had been successfully introduced into the exon 1 *B2M* locus. However, the PLB-E6₂₉₋₃₈-(GGGGS)₃ protein was apparently not synthesized by gene-edited cells, in contrast to when introduced as an independent sequence expressed in a plasmid format (Figure 3.13). Both plasmid and gene-edited formats of PLB- E6₂₉₋₃₈-(GGGGS)₃ resulted in reduced MHC-I cell surface expression (Figure 3.12 and 4.9). In this section, I chose to evaluate the use of PLB N₃₂₅₋₃₃₃ of SARS-CoV-2 peptide and the (GGGGS)₃ linker, which demonstrated a high-level of expression and rescue of MHC-I (Figure 3.13).

Firstly, N₃₂₅₋₃₃₃-(GGGGS)₃ sequence was knocked-in into B2M.exon1 of HEK cells by co-transfection of Cas9 RNP of sgRNA.B2M.1.3 and Megamer-N₃₂₅₋₃₃₃-(GGGGS)₃ using the Amaxa Nucleofector system. Genomic DNAs were extracted from one day post-transfected cells and used for preparing amplicons for the T7EI assay by F1.B2M.exon1 and R1.B2M.exon1 primers. Electrophoresis analysis demonstrated that only samples which were transfected with both Cas9 RNP and PLB donor template had both amplicons of wt *B2M* locus (762 bp) and the HDR band (834bp) containing the B2M gene with the inserted N₃₂₅₋₃₃₃ peptide and (GGGGS)₃ linker sequences (Figure 4.11).

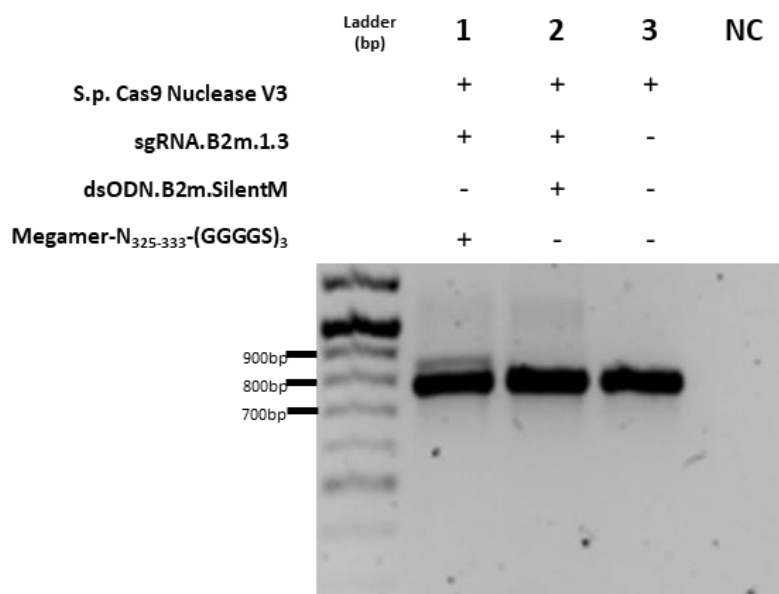


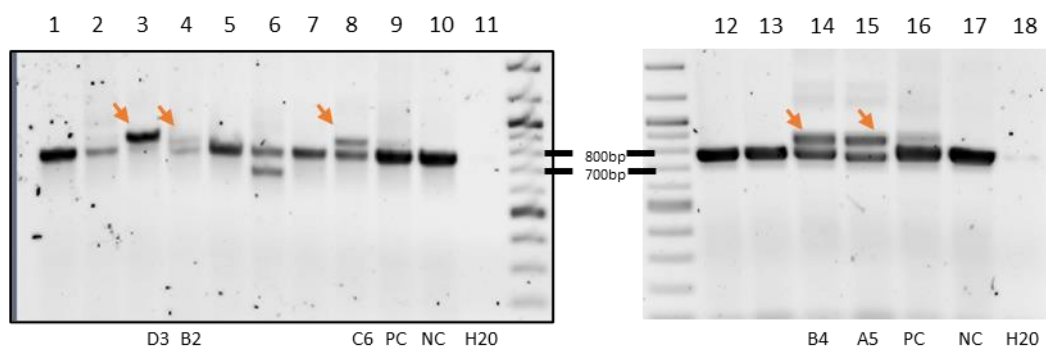
Figure 4.11. HDR results of PLB- N₃₂₅₋₃₃₃-(GGGGS)₃ insertion in HEK cells

Megamer-N₃₂₅₋₃₃₃-(GGGGS)₃ donor templates and RNP.B2M-1.3 were co-transfected into HEK cells by electroporation. At one day post-transfection, PCR amplicons of transfected cell-derived genomic DNA were used to check HDR efficiency by PCR. The size of wt B2M and HDR amplicons were 762 and 834 bp, respectively. Data are presented of one experiment.

Monoclonal HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ clones were isolated as previously described and partly harvested for extraction of genomic DNA. Clones were screened for PLB specific amplicons. 5 in 25 (20%) screened cell clones exhibited amplicons correlating to the insertion of the N₃₂₅₋₃₃₃-(GGGGS)₃ sequence within *B2M* (Figure 4.12A). Amongst the clones, 4 were potentially heterozygous for the HDR gene-edited HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ sequence i.e., B2, -C6, -B4, and -A5 (lane 4, 8, 14 and 15). Only HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ clone D3 (lane 3) was potentially a homozygous HDR clone with N₃₂₅₋₃₃₃-(GGGGS)₃ sequence inserted into both *B2M* copies. The PCR product of exon 1 of the *B2M* locus of HEK-N₃₂₅₋₃₃₃-(GGGGS)₃-D3 clone was sequenced using the F1.B2M.exon1 primer. Sanger sequencing (Figure 4.12B) revealed that the

sequence of exon 1 *B2M* locus of HEK-N₃₂₅₋₃₃₃/(GGGGS)₃-D3 had indeed the introduced Megamer-N₃₂₅₋₃₃₃-(GGGGS)₃ sequence without any additional indel mutations. Therefore, the above data suggests that N₃₂₅₋₃₃₃ peptide and (GGGGS)₃ linkage was successfully inserted into the end of exon 1 of both *B2M* gene copies within HEK-N₃₂₅₋₃₃₃/(GGGGS)₃-D3 and was therefore homozygous for the PLB sequence (Figure 4.12B).

A



B

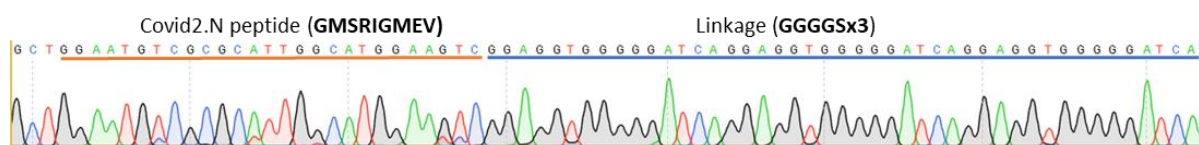


Figure 4.12. Single-cell isolation for monoclonal HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ clones from polyclonal gene-edited populations.

(A) Monoclonal gene-edited cells were isolated using serial dilution of polyclonal populations followed by DNA extraction for PCR with F1.B2M.exon1 and R1.B2M.exon1 primers. Wild-type and HDR bands have the length of about 772 bp and 834 bp, respectively. Electrophoresis analysis of PCR screening was used to identify HEK clones containing the HDR of N₃₂₅₋₃₃₃/(GGGGS)₃. (B) Insertion of N₃₂₅₋₃₃₃ SARS-CoV-2 peptide and GGGGS₃ linker was confirmed by Sanger sequencing of B2M.exon1 in HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ clone D3. Data are presented of one experiment.

To evaluate expression of surface MHC-I by gene-edited cells, equal numbers of each HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ clones were analysed by flow-cytometry using W6/32, BB7,1 (anti- *HLA-B*07:02*) and BB7.2 antibodies [342, 350]. HEK and HEK.B2M.KO were used as positive and negative controls for expression of surface MHC-I. Before antibody staining, un-transfected and gene-edited clones were incubated with IFN- γ (400 IU/ml) for 3 days to enhance MHC-I expression. Flow-cytometry data of each sample stained by each antibody were compared to control HEK samples using ordinary one-way ANOVA (Prism) (Appendix 4, Supplementary Data 8, 9 and 10). Flow cytometric analysis (Figure 4.13) revealed that W6/32- and BB7.1 reactive HLA types were expressed at lower levels by all HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ cells compared to wild type HEK cells. While HLA-A2 expression levels of heterozygous HDR HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ clone B2, C6 and A5 (MFI:15222 \pm 876, 16124 \pm 499, 15441 \pm 58) were not significantly different when compared with WT HEK cells (MFI:17322 \pm 414), those of the homozygous HDR HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ clone D3 (MFI: 12345 \pm 1104) and heterozygous HDR HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ clone B4 (MFI: 2939 \pm 58) were significantly lower than wt HEK (Figure 4.13B). The flow cytometry analysis suggests that PLB of N₃₂₅₋₃₃₃/(GGGGS)₃ potentially reduces the expression of MHC-I molecules in general but possibly not the peptide-specific HLA class I complexes. However, this effect was not observed amongst all monoclonal gene-edited HEK cells. Therefore, the next stage, involved determining the protein expression of PLB by immunoblotting.

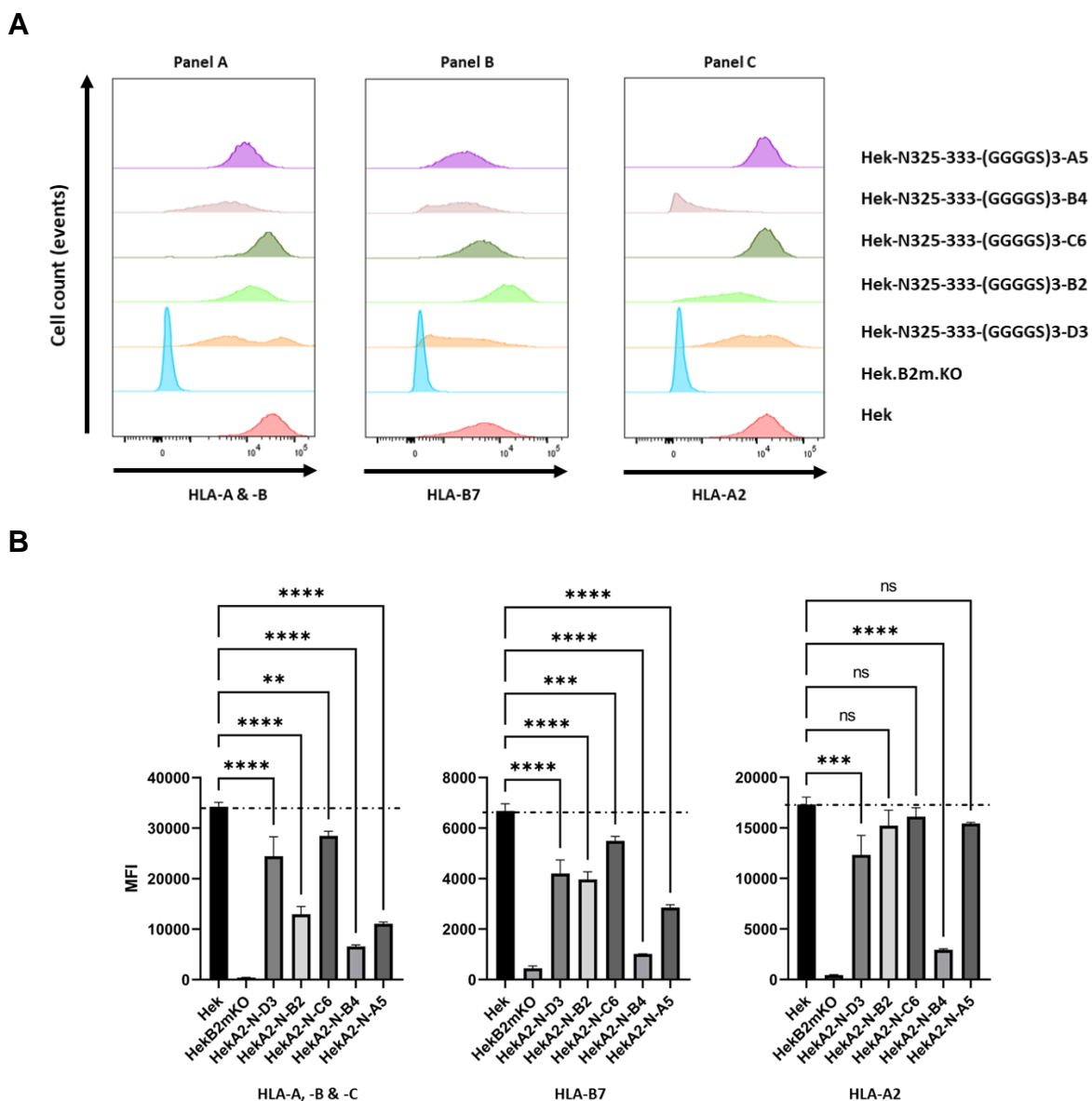


Figure 4.13. Characterization of surface MHC-I expression on HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ cells.

(A) Cell surface expression of MHC-I (panel A), HLA-B7 (Panel B) and HLA-A2 (panel C) by HEK, B2M-deficient cells and monoclonal HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ cells were analyzed by flow cytometry using W6/32, BB7.1 and BB7.2 antibodies (Top panel). Histogram plots of cell count versus MFI corresponding to the surface expression of MHC-I, HLA-B7 and HLA-A2 were presented. (B) Bar charts of the mean MFI data from cell surface expression of MHC-I on monoclonal HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ cells. The mean MFI is calculated from triplicated experiments (mean ± SEM, n=3). One-way ANOVA with Turkey's post-test. ****P < 0.001.

Equal number (10^6) of each monoclonal HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ cell line were lysed in 1% NP-40 detergent. Reduced lysates were resolved on 17% SDS-PAGE, followed by immunoblotting with anti-B2M and anti-GAPDH antibodies (Figure 4.14, bottom blot) with cell lysates from wild type HEK (lane 1) and HEK.B2M.KO (lane 2) cells used as positive and negative controls for expression of wild type B2M (11.7 kDa) protein, respectively (Figure 4.14, top blot).

Cell lysates generated from HEK.B2M.KO cells transfected with *pCR3.1-N₃₂₅₋₃₃₃-(GGGGS)₃-B2M* were used as a positive control for expression of PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ (13.6 kDa, lane 9) (previously described in section 3.3.5). Immunoblotting revealed the expression a band of ~11.7 kDa in size in all five monoclonal HEK- N₃₂₅₋₃₃₃-(GGGGS)₃ cells (lane 3-7), suggesting only wt B2M protein was being expressed and not the PLB protein.

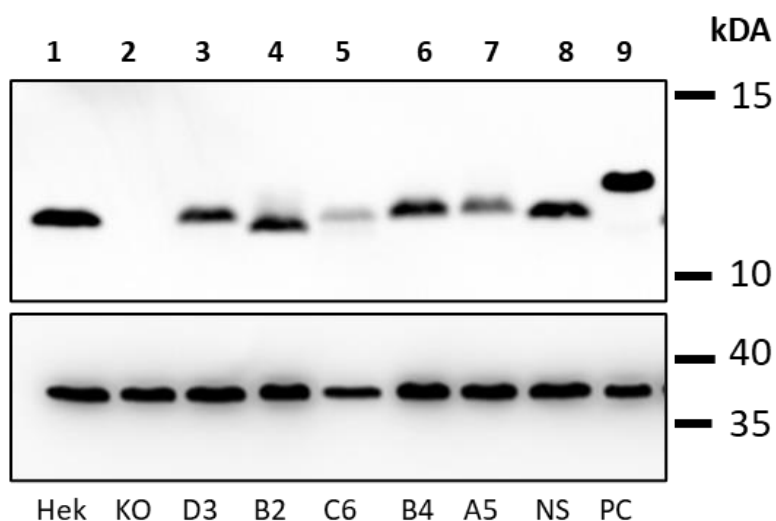


Figure 4.14. Immunoblotting of B2M protein on HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ cells.

HEK, HEK.B2M.KO and HEK-N₃₂₅₋₃₃₃/(GGGGS)₃ cells were lysed in 1% NP-40 detergent. Reduced lysates were resolved on 17% SDS-PAGE followed by immunoblotting with anti-B2M and anti-GAPDH antibodies. Cell lysate of HEK.B2M.KO transfected with *pCR3.1-N₃₂₅₋₃₃₃-(GGGGS)₃-B2M* was used as positive control for expression of PLB-N₃₂₅₋₃₃₃-(GGGGS)₃.

(GGGGS)₃ (lane 8). The expected size of B2M and PLB-N₃₂₅₋₃₃₃/(GGGGS)₃ proteins are 11.7 and 13.6 kDa, respectively. Data are presented of one experiment.

It would appear that cells can synthesis PLB-N₃₂₅₋₃₃₃–(GGGGS)₃ when delivered in a pPLB plasmid format but not as an introduced gene-edited sequence. The PLB plasmid construct generated the protein of correct size, so the lack of PLB expression was unlikely to be a post-translational event. The main difference in approaches were – the plasmid was introducing the PLB as a non-*B2M* integrated sequence, while gene editing was introducing the sequence into the *B2M* gene. Therefore, it was possible that the introduction of additional sequences within the *B2M* locus close to the exon-intron boundary could affect transcriptional processing. Therefore, I examined the sequence and identified potential sequences which could interfere with intron/exon processing.

Based on the genomic sequence, intron 1 of the *B2M* gene has the intron splicing motifs of GU at the 5' end and AG at the 3' end (top image, Figure 4.15). GU motifs were also detected in the sequence of the N₃₂₅₋₃₃₃ peptide and (GGGGS)₃ linker (bottom image, Figure 4.15) DNA sequences. This preliminary sequence analysis suggested that insertion of N₃₂₅₋₃₃₃–(GGGGS)₃ was introducing potential GU motifs into intron 1 of the *B2M* locus. As a result, the peptide-linker sequences could therefore be removed as intronic elements to form different mature mRNAs leading to the translation of WT β2m protein or β2m exhibiting various sizes but not containing the complete PLB protein sequence as was evidenced in Figure 4.15.

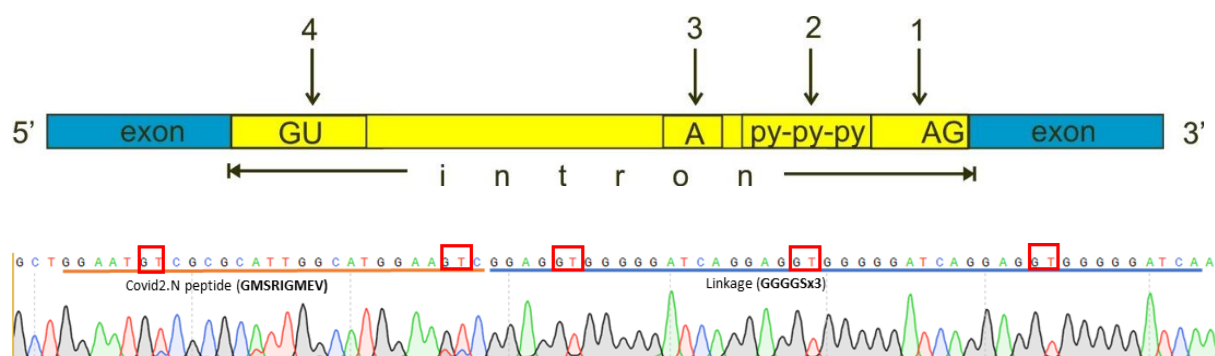


Figure 4.15. Illustration of intron motif on $N_{325-333}-(GGGGS)_3$.

Top image illustrates exons and introns motif in pre-mRNA. The bottom image indicates the intron motif (red squares) were identified on $N_{325-333}-(GGGGS)_3$ sequence.

To test the above, I initially examined mRNA expression by extracting total mRNAs used for cDNA synthesis by reverse transcriptase using Oligo(dT) primers. Synthesised cDNA was then used as templates for PCR using F1.B2M.cDNA and R1.B2M.cDNA to detect the existence of B2M and PLB cDNA. The expected PCR amplicon sizes of B2M and PLB- $N_{325-333}-(GGGGS)_3$ were 281 and 353 bp, respectively. HEK-derived mRNA was used as positive control for the amplification of wild-type B2m cDNA (lane 1). The amplification of wild-type B2m cDNA from HEK-derived mRNA without Reverse transcriptase was used as negative control for the RT reaction. RT (+) HEK.B2M.KO was used as a negative control of B2M mRNA (lane 2). *pCR3.1-B2M* and *pCR3.1-N₃₂₅₋₃₃₃-(GGGGS)₃-B2M* plasmids were used as PCR positive controls for B2M and PLB sequences. PCR products from HEK- $N_{325-333}-(GGGGS)_3$ clones D3, B2, C6 and B4 cDNAs (lane 3-6) (Figure 4.16) revealed smear bands with size ranges of approximately 300-350 bp. The smear bands ranged from the expected size of PLB- $N_{325-333}-(GGGGS)_3$ of approximately 350bp and to that of B2M of approximately 300bp. These data are strongly indicative that $N_{325-333}$ peptide

and (GGGGS)₃ linker sequences have been potentially removed from the mRNA of PLB using alternate intronic splice and acceptor sites which were introduced as part of the PLB sequence.

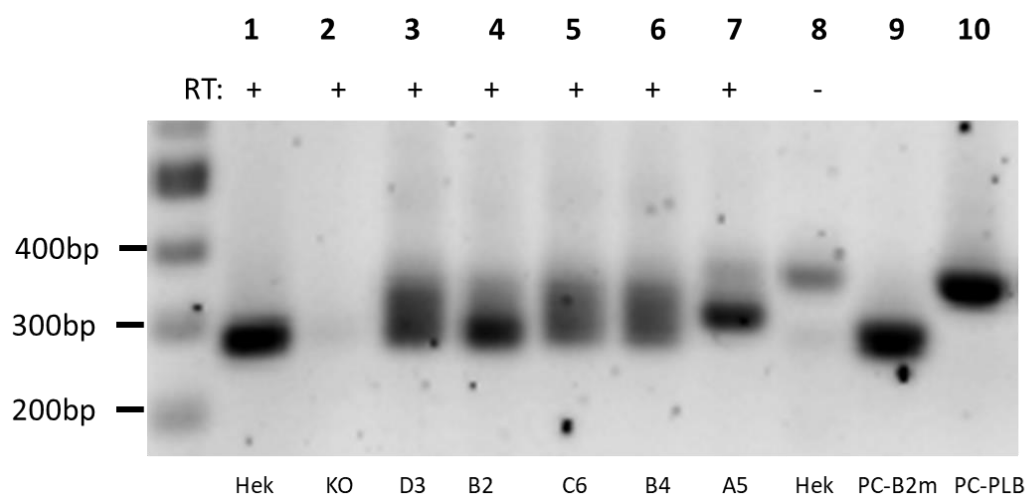


Figure 4.16. Detection of PLB mRNA expressed by gene-edited cells.

Total mRNAs were extracted from HEK, HEK.B2M.KO and HEK- N₃₂₅₋₃₃₃-(GGGGS)₃ and used for cDNA synthesis. cDNA templates were used as templates for PCR with primers specified for B2M. The expected PCR amplicon sizes of B2M and PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ are 281 and 353 bp, respectively. *pCR3.1-B2M* and *pCR3.1-N₃₂₅₋₃₃₃-(GGGGS)₃-B2M* plasmids were used as a PCR positive control for B2M and PLB. Data are presented of one experiment.

In summary, the N₃₂₅₋₃₃₃ peptide and (GGGGS)₃ linker have been successfully introduced into exon 1 of the *B2M* locus of HEK cells using the CRISPR/Cas9 system and Megamer donor template. By single-cell cloning, 5 monoclonal HEK-N₃₂₅₋₃₃₃-(GGGGS)₃ cells, including 1 homozygous and 4 heterozygous populations were isolated and expanded. However, due to the potential introduction of intron acceptor/donor splicing motifs within the N₃₂₅₋₃₃₃ peptide and (GGGGS)₃ linker, the mRNA sequences of these components were possibly recognized as an intronic area

of *B2M* and therefore cleaved during the intron splicing process. As a result, PLB proteins were not expressed or expressed as shorter versions without the desired peptide and linker. Consideration of introducing intron motifs for gene-editing purposes have not been previously reported to the best of my knowledge, by any previous publication. While some intron motif sequences within the linker could be easily replaced, some within the N₃₂₅₋₃₃₃ peptides were unchangeable due to the risk of introducing missense mutations. Although I was not able to repeat the experiment with the N₃₂₅₋₃₃₃ peptide sequences, the motif removal has been a useful lesson for the insertion of other peptides.

4.2.5 Expression of PLB-S₂₆₉₋₂₇₇-G₆(SGG)₃ by HEK cells

In this section, SARS-CoV-2 S₂₆₉₋₂₇₇ peptide and G₆(SGG)₃ linker were the next candidates for generating PLB molecules by CRSIPR/Cas9 technology. Previously, pCR3.1-S₂₆₉₋₂₇₇ plasmids had shown an ability to rescue expression of HLA-A2 heavy chains (section 3.3.5). Especially, the use of the G₆(SGG)₃ linker enhanced the expression of surface HLA-A2 by four times compared to the (GGGGS)₃ linker. In addition, the removal of intronic motifs were also considered in this experiment to overcome the issue with previous peptides and maximize the chance of PLB synthesis. All intronic motifs within the peptide S₂₆₉₋₂₇₇ and linker G₆(SGG)₃ sequence were altered by introducing silent base pair changes; for example GT to GA (Figure 4.17).



Figure 4.17. Removal of potential intronic splice/acceptor motifs within the G₆(SGG)₃ linker.

Image illustrates protein sequence of the G₆(SGG)₃ linker (brown) and S₂₆₉₋₂₇₇ peptide (green). Nucleotide substitutions (yellow highlighted) were included in the DNA template sequence of S₂₆₉₋₂₇₇-G₆(SGG)₃ to replace all possible intronic motifs by silent mutations.

Cas9 RNP of sgRNA.B2M.1.3 and Megamer-S₂₆₉₋₂₇₇-G₆(SGG)₃ were co-transfected into HEK cells using electroporation with the Amaxa Nucleofector system and assayed for HDR by PCR using F1.B2M.exon1 and R1.B2M.exon1 primers followed by T7EI assay. HEK cells co-transfected with the Megamer-N₃₂₅₋₃₃₃-(GGGGS)₃ and RNP.B2M.1.3 were used as a positive control for HDR of the PLB. The T7EI assay

(top image, Figure 4.18) revealed cleaved bands of 458 and 304 bp could be generated from HEK cells transfected with Cas9+sgRNA.B2M.1.3 RNP (lane 1, 2, and 3), indicating that Cas9/sgRNA.B2M.1.3 RNP complexes can introduce DSBs and lead to indel mutations (without donor template) or with the desired insertion (with donor template). PCR products (top image, Figure 4.18) of 834 bp correlating to the HDR sequence were observed in HEK transfected with RNP and the donor template (lane 1 and 2), therefore indicating the sequence of S₂₆₉₋₂₇₇-G₆(SGG)₃ has possibly been inserted into exon 1 of the *B2M* locus in HEK cells.

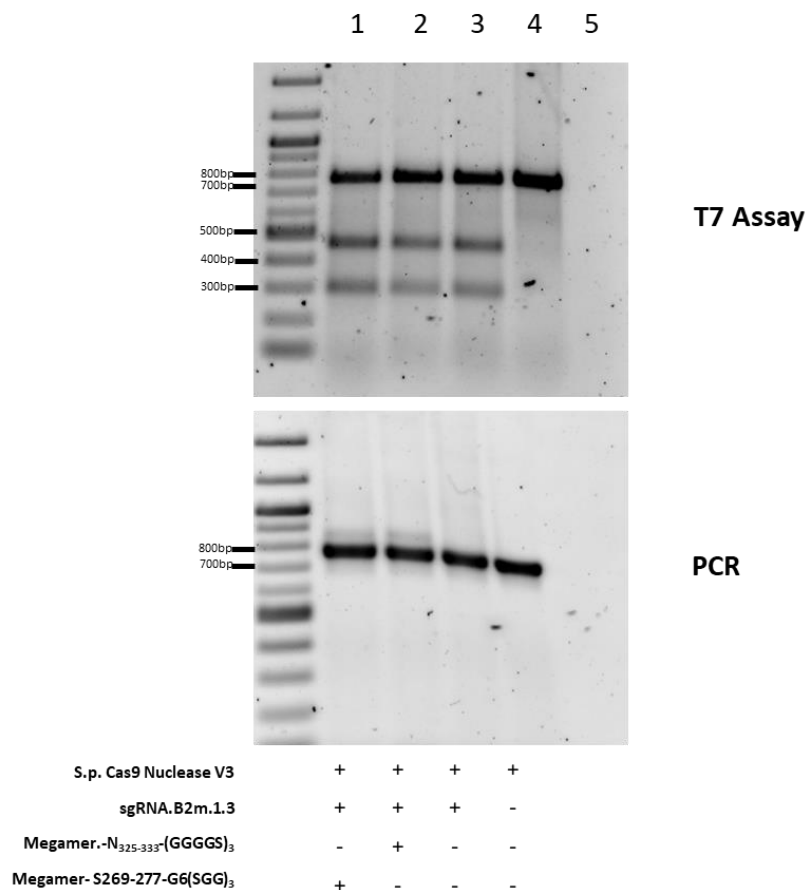


Figure 4.18. Results from the HDR of PLB- S₂₆₉₋₂₇₇-G₆(SGG)₃ sequence insertion into HEK cells.

Megamer- S₂₆₉₋₂₇₇ -G₆(SGG)₃ and RNP.B2M-1.3 were co-transfected into HEK cells. At one day post-transfection, PCR amplicons of transfected cell-derived genomic DNA were used to determine HDR efficiency (bottom gel) and T7EI assay (top gel). The sizes of the wt B2M and HDR amplicon were 762 and 834 bp, respectively. Data are presented of one experiment.

Single cell clones were selected, and genomic DNA was extracted. Clones were screened for DNA template insertion using PCR screening with F1.B2M.exon1 and R1.B2M.exon1 primers. 4 in 58 (6.7%) screened cell clones potentially contained the S₂₆₉₋₂₇₇/G₆(SGG)₃ sequence inserted within the B2M genomic sequence (data not shown). PCR screening revealed products from 3 HEK-S₂₆₉₋₂₇₇ -G₆(SGG)₃ clones - 3D1, 2B1, and 2C6 which contained the PLB-like bands (834 bp) and wt B2M-like band (762 bp) (Figure 4.19). The HEK-S₂₆₉₋₂₇₇ -G₆(SGG)₃-3A5 clone exhibited an upper band of higher than the expected Mw of the PLB at around 900 bp. Therefore, all 4 HEK-S₂₆₉₋₂₇₇ -G₆(SGG)₃ clones were heterozygous for B2M but only HEK-S₂₆₉₋₂₇₇ -G₆(SGG)₃ -1D3, 2B1, and 2C6 clones possibly contained the desired HDR modifications.

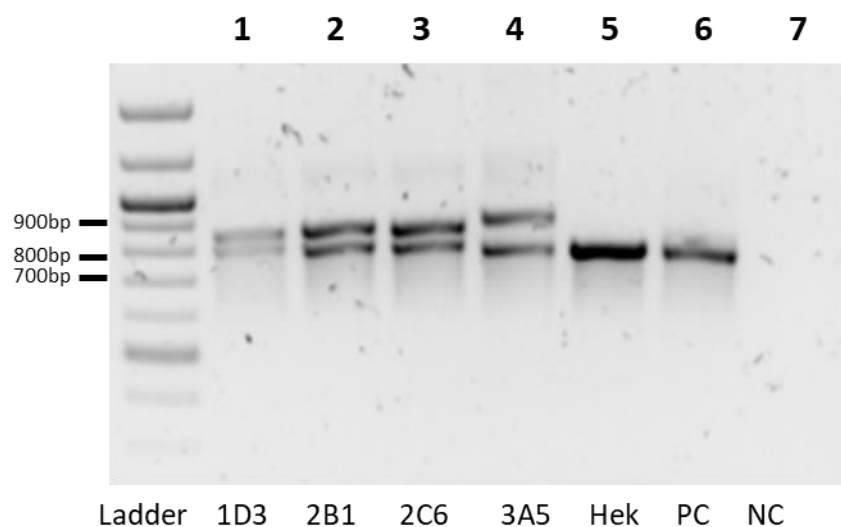


Figure 4.19. Single-cell HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ clones assessed for HDR insertion of S₂₆₉₋₂₇₇-G₆(SGG)₃ sequence.

Monoclonal gene-edited cells were isolated using serial dilution of polyclonal populations followed by DNA extraction for PCR with F1.B2M.exon1 and R1.B2M.exon1 primers. Wild-type and HDR bands have the length of 772 bp and 834 bp, respectively. Electrophoresis analysis of PCR screening used to identify HEK clones containing the HDR of S₂₆₉₋₂₇₇-G₆(SGG)₃. HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ samples were used as negative and positive controls for HDR of PLB. Data are presented of one experiment.

To evaluate expression of surface MHC-I on gene-edited cells, equal cell numbers of HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ clones were analysed by flow-cytometry using W6/32, BB7.1 and BB7.2 antibodies. HEK and HEK.B2M.KO was used as positive and negative control for expression of surface MHC-I followed by incubation with IFN- γ (400 IU/mL) for 3 days to enhance expression of MHC-I. Flow-cytometry data of each sample stained by each antibody were compared to control HEK samples using ordinary one-way ANOVA (Prism) (Appendix 4, Supplementary Data 11, 12 and 13). Flow cytometric analysis (Figure 4.20) revealed that clones 2C6 and 1D3 expressed comparable levels of MHC-I by WT HEK cells as determined by W6/32 staining. However, HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-3A5 clone expressed significantly lower levels than that of wild type HEK cells. However, compared to wt HEK cells, HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-1D3 clone expressed significantly higher MFI of MHC-I, HLA-B7 and HLA-A2. While HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-2B1 had the lower expression level of MHC-I in general as detected by W6/32, that of HLA-A2 and HLA-B7 were not significantly different compared to WT HEK cells. It is possible that the PLB protein HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-2B1 may not affect expression of HLA-A2 and HLA-B7 but can affect other MHC-I molecules expressed by HEK cells. In contrast, HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-2C6

had the higher expression level of HLA-A2 in general but that of MHC-I and HLA-B7 were not significantly different compared to WT HEK cells, suggesting that the PLB proteins of HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-2B1 may maintain the expression of MHC-I and enhance expression of HLA-A2. In summary, flow cytometric analysis revealed that different HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ clones had different impacts on surface expression of HLA class I. Therefore, investigation of protein production was required to understand what form of PLB has been synthesised by gene-edited cells.

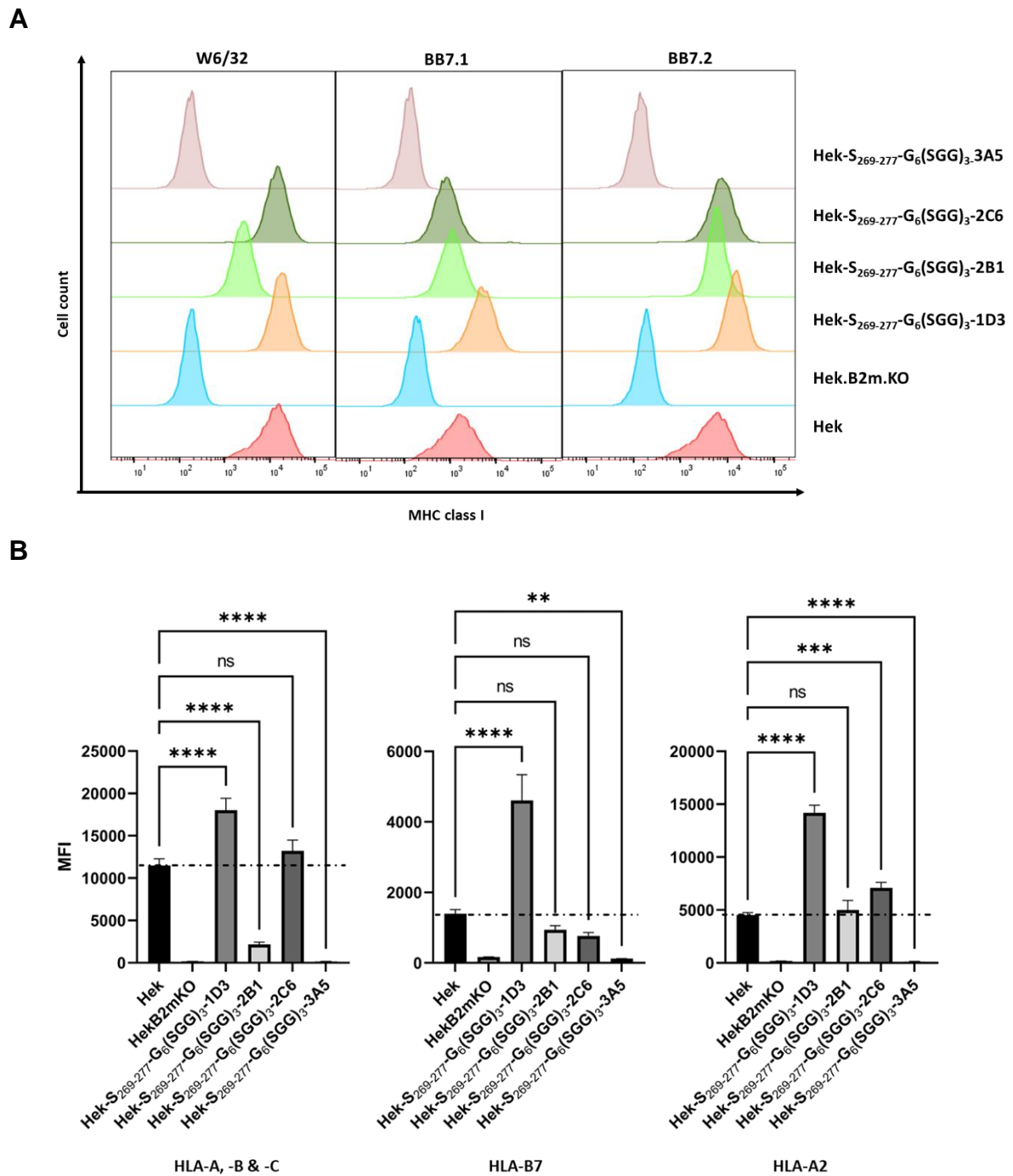


Figure 4.20. Flow-cytometry analysis of MHC-I expression by HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells.

(A) The flow cytometry histograms of surface MHC-I, HLA-B7 and HLA-A2 expression by HEK, B2M-deficient HEK and monoclonal HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells. Surface MHC-I,

HLA-B7 and HLA-A2 expression was determined by flow cytometry using W6/32, BB7.1 and BB7.2 antibodies, respectively. (B) The bar charts indicate the Mean Fluorescence Intensities (MFI) of MHC-I expression. One-way ANOVA with Turkey's post-test. **** $P < 0.001$. The means of MFI were calculated from triplicate data (mean \pm SEM, $n=3$).

To examine the expression of B2M and PLB proteins by HEK- $S_{269-277}\text{-G}_6(\text{SGG})_3$ clones, equal numbers (10^6) of cells were lysed in 1% NP-40 detergent. Furthermore, the culture supernatant was also collected to determine the presence of free B2M proteins, which could be released from the recycling of MHC-I at the plasma membrane [351]. The detection of cell-free B2M could indicate the effective transportation of MHC-I containing wt B2M as well as PLB from ER to cell membrane. Reduced lysates and supernatant were resolved on 17% SDS-PAGE followed by immunoblotting with anti-B2M and anti-GAPDH antibodies (Figure 4.21). HEK (lane 1 and 8) and HEK.B2M.KO cell lysates (lane 2 and 9) were used as positive and negative controls for B2M expression. Cell lysates generated from HEK.B2M.KO transfected with *pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M* was used as a positive control for expression of PLB- $S_{269-277}\text{-G}_6(\text{SGG})_3$ with the size of 13.6 kDa (lane 7) (previously described in section 3.3.5).

Immunoblotting revealed that there were no B2M or PLB bands detected in both cell lysate and growth medium of HEK- $S_{269-277}\text{-G}_6(\text{SGG})_3\text{-3A5}$ cells (lane 6 and 13, respectively). Only B2M-like protein band was observed in cell lysates of HEK- $S_{269-277}\text{-G}_6(\text{SGG})_3\text{-2C6}$ clone (lane 5) and only PLB-like protein band were detected in both cell lysate and growth medium of HEK- $S_{269-277}\text{-G}_6(\text{SGG})_3\text{-2B1}$ suggesting that PLB- $S_{269-277}\text{-G}_6(\text{SGG})_3$ was being synthesised and binding to cell surface MHC-I complexes. Both B2M- and PLB- like bands were observed in both cell lysate and

growth medium of HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-1D3 (lane 3 and 10, respectively), indicating that the B2M alleles of HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃-1D3 was able to synthesis the functional light chains which can form the complex with MHC-I and be transported to the cell surface.

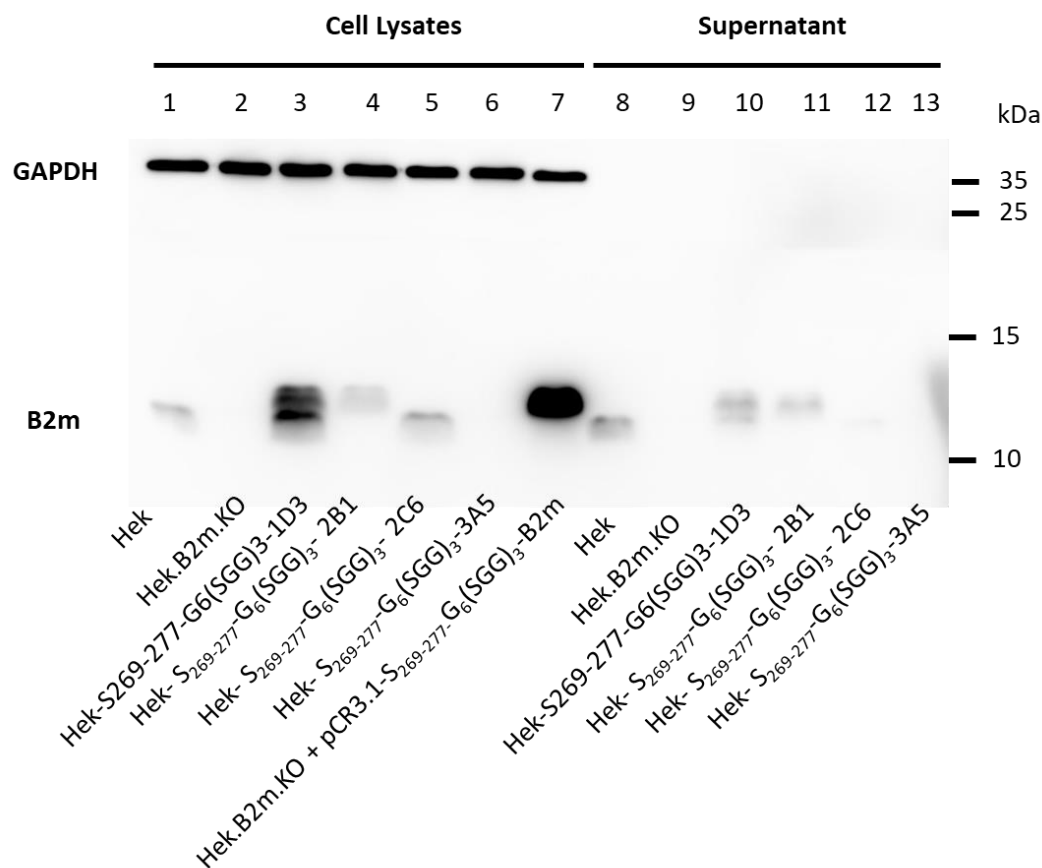


Figure 4.21. Immunoblotting of B2M in HEK- S₂₆₉₋₂₇₇-G₆(SGG)₃ clonal cell populations.

Cell growth medium and cell lysates in 1% NP-40 detergent were collected from monoclonal HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells and resolved by 17% SDS-PAGE. Separated proteins were immunoblotted with anti-B2M and anti-GAPDH antibodies. HEK and HEK.B2M.KO cell lysates were used as positive and negative controls for B2M protein expression. Cell lysates of HEK.B2M.KO cells transfected with *pCR3.1-N₃₂₅₋₃₃₃-(GGGS)₃-B2M* was used as a positive control for PLB expression. The expected size of B2M and PLB- S₂₆₉₋₂₇₇-G₆(SGG)₃ proteins were 11.7 and 13.6 kDa, respectively. Data are presented of one experiment.

In summary, the S₂₆₉₋₂₇₇ peptide and G₆(SGG)₃ linker were re-designed to remove all intronic motifs and have been successfully introduced into exon 1 of the *B2M* locus of HEK cells using the CRISPR/Cas9 system and a Megamer donor template. By single-cell cloning, 4 monoclonal *B2M* heterozygous HEK- S₂₆₉₋₂₇₇-G₆(SGG)₃ cells have been isolated and expanded. While HEK- S₂₆₉₋₂₇₇-G₆(SGG)₃-3A5 cells did not synthesis either PLB or B2M protein which led to the absence of surface MHC-I, HEK- S₂₆₉₋₂₇₇-G₆(SGG)₃-3A5 only expressed B2M-like protein but not the PLB. Among 4 monoclonal cell lines, HEK- S₂₆₉₋₂₇₇-G₆(SGG)₃ clones 1D3 and 2B1 have shown to produce PLB-like proteins which can bind to HLA class I and form MHC-I- peptide complexes as implied by W6/32 and BB7.2 antibody staining. Flow cytometry analysis and immunoblotting on cell growth medium, suggests that the MHC-I molecules were able to be appear at the cell surface on HEK- S₂₆₉₋₂₇₇-G₆(SGG)₃ -1D3 and -2B1 cells.

4.2.6 Evaluate off-target effects on gene-edited cell lines

CRISPR-mediated gene editing technology can introduce undesirable off-target mutations at genome regions which have high level of homology to the targeted sequence [352]. Therefore, deep sequencing was applied to evaluate the off-target effects caused by the Cas9 RNP of sgRNA.B2M.1.3 for safety concerns. Firstly, potential off-target sequences were predicted using COSMID, an online software tool for identifying and validating CRISPR Off-target regions [353]. The criteria for identifying off-target sites included: (i) screening on whole human genomes; (ii) containing maximum of 3 mismatch and 1 indel mutation; (iii) <2.5 in score of off-target sites for the input target sites. Lower ranking scores of off-target sites indicated the more likely off-target sites. Following the above criteria, I identified 1 on-target site (OT.sgB2M1.3_1) and 2 potential off-target sites (OT.sgB2M1.3_2 and OT.sgB2M1.3_3) (Table 4.1). While the OT.sgB2M1.3_1 is the targeted site of B2M, OT.sgB2M1.3_2 and OT.sgB2M1.3_3 are intergenic regions located on chromosome 13 and chromosome 20, respectively.

Table 4.1. On-target and Off-target sites for RNP of Cas9 protein and sgRNA.B2M.1.3.

Off-target regions of sgRNA.B2M.1.3 were predicted by COSMID software. Mismatch or indel mutations are highlighted in yellow. Query type indicates the type and number of changes examined. Off-target ranking scores are based on the number and location of base mismatches.

Name	Result	Query type	Mismatch	Chr Position	Strand	Score
OT.sgB2M1.3_1	ACTCACGCTGGATAGCCTCCAGG -- hit ACTCACGCTGGATAGCCTCCNGG -- query	No indel	0	Chr15:44711597- 44711619	-	0
OT.sgB2M1.3_2	ACTCACCCCTGGTAGCCTCCAGG -- hit ACTCACGCTGGTAGCCTCCNGG -- query	Del 9	1	Chr13:53068362- 53068383	-	1.54
OT.sgB2M1.3_2	ACTCACTGCTGGACAGCCTCCTGG -- hit ACTCACNGCTGGATAGCCTCCNGG -- query	Ins 14	1	Chr20:22332541- 22332564	+	2.03

Primers for NGS were designed to produce short PCR amplicons with sizes of 200 to 250 bp (Appendix 5). Genomic DNA of wt HEK cells and polyclonal gene-edited HEK cells, which were transfected with RNP of Cas9 protein and sgRNA.B2M.1.3, were extracted and used as the template for preparing the initial PCR amplicons. With ~280 times higher fidelity amplification than *Taq* polymerase, Q5 DNA Polymerase (NEB) was used to minimize the error rate during the process of amplification. After generating libraries with unique bar codes for each sample, PCR products were subsequently sequenced on the Illumina Miseq at 100,000 reads per sample.

Off-target effects were evaluated based on the comparison between NGS sequencing data and reference genome sequences. The percentage of variants or indel mutations were analyzed using re-alignment programs Geneious using the following criteria: (i) identifying variants within the entire range covered by the sequencing reads; (ii) eliminating variants with low frequency (<0.5%). Off-target analysis (Figure 4.22) indicates that on-target sites contained nearly 30% indel mutations generated during DSB repair following the NHEJ pathway. More importantly, the percentage indels detected in 2 off-target regions of gene-edited cells were very low at 0.1% for OT.sgB2M1.3_2 and 0.3% for OT.sgB2M1.3_3. These percentages and types of variants were identical to that of untreated HEK cells. Therefore, off-target analysis indicated that indels were not detected in two screened off-target loci of sgRNA.B2M.1.3.

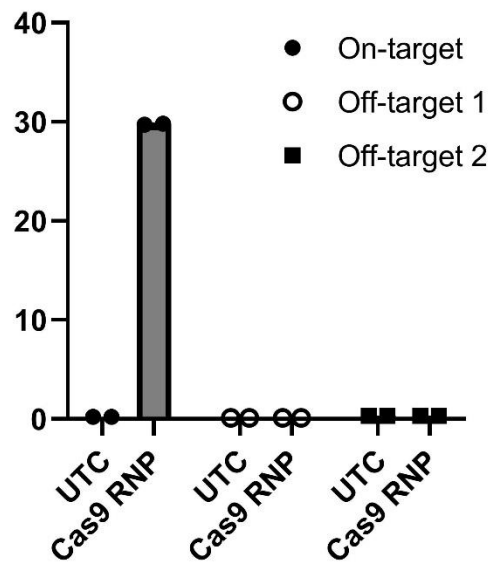


Figure 4.22. On- and Off-target analysis of HEK cells treated with Cas9 RNP with sgRNA.B2M.1.3.

Indel frequencies for on-target and potential off-targeted sequences identified by Miseq deep-sequencing in untreated HEK cells (n=2) and CRISPR/Cas9-treated HEK cells (n=2).

4.3 Discussion

In this chapter, I have developed an efficient CRISPR/Cas9-mediated system that precisely inserts the peptide and linker into the human *B2M* locus. The modification facilitates the production of Peptide-linked B2M (PLB) molecules which rescue the expression of peptide-specific HLA class I at the cell surface. The MHC complexes of PLB and peptide-specific HLA are expected to present desired peptides for inducing recognition by CTLs.

The E6₂₉₋₃₈ peptide was previously described to be restricted by HLA-A*02:01 and reported to be the endogenously processed T-cell epitope of HPV 16 E6 protein [338]. Therefore, I used the glycine rich (GGGGS)₃ sequence to link the HLA-A2 restricted HPV 16 E6 epitope (E6₂₉₋₃₈) to human B2M to test its assembly with HLA class I heavy chain. Although the E6₂₉₋₃₈ peptide and (GGGGS)₃ linker were successfully knocked-into exon 1 of *B2M*, there was no PLB expression detected in monoclonal HEK- E6₂₉₋₃₈-(GGGGS)₃, therefore resulting in the absence of cell surface MHC-I.

The experiment was repeated using SARS-CoV-2 N₃₂₅₋₃₃₃ peptide which has shown to rescue HLA-A2 expression in B2M-deficient HEK cells transfected with the pCR3.1-N₃₂₅₋₃₃₃-(GGGGS)₃-*B2M* plasmid. The expression of both peptide specific and non-specific HLA class I binding molecules were detected at different levels in 5 monoclonal Hek-N₃₂₅₋₃₃₃-(GGGGS)₃ cells. Immunoblotting showed that cell lysates of all Hek-N₃₂₅₋₃₃₃-(GGGGS)₃ clones only contained B2M-like proteins but not PLB molecules. I hypothesized that the peptide and linker sequences were potentially cleaved from PLB mRNA before translation. The smear bands detected by B2M PCR products of RNA isolated from clones confirmed several PLB mRNA species existed in Hek-N₃₂₅₋₃₃₃-(GGGGS)₃ clones. The mRNA variants can be explained by the impact

of the intronic removal process which were cleaved at conserved sequences during the intron-exon splicing process. Indeed, five dinucleotide GUs, with common intronic motifs at the 5' end within the N₃₂₅₋₃₃₃ peptide and (GGGS)₃ linker were identified. These newly inserted GU sequences could become the location where intronic removal begins, and end with AG at its 3' end. The different number of GU sequences could therefore produce heterogeneous mRNA populations. Depending on the location of GU motifs, these could be removed by merely changing the codon usage, however if located within the peptide sequence and depending on the sequence, it may not be possible to alter the nucleotides to erase these motif sequences without changing amino acid sequences. Without appropriate T cells, we could not assess how potential peptide changes could impact on T cell recognition. However, this was the very first study investigating the impact of intron removal during pre-mRNA splicing on CRISPR-mediated nucleotide modifications.

Previously I demonstrated that transfection of pCR3.1-S₂₆₉₋₂₇₇-G₆(SGG)₃-B2M plasmid could enhance the expression of HLA-A2 on the cell surface of B2M-deficient HEK cells at much higher levels compared to other PLB plasmids. Additionally, using the NETMHCpan algorithm prediction tool, S₂₆₉₋₂₇₇ peptide was postulated to strongly bind to HLA-A2 compared to a set of random natural peptides. To-date, SARS-CoV-2 Spike S₂₆₉₋₂₇₇ appears to be the most immunodominant SARS-CoV-2 epitope found in individuals bearing the HLA-A*02:01 allele [354]. Therefore, S₂₆₉₋₂₇₇ peptide was chosen as the next candidate to test the HDR of PLB in HEK cells. To maximize expression of HLA-A2, the G₆(SGG)₃ linker was used. Furthermore, the intronic motifs within the S₂₆₉₋₂₇₇-G₆(SGG)₃ donor template sequence was replaced by silent mutations. The S₂₆₉₋₂₇₇ peptide and G₆(SGG)₃ linker were successfully knocked-in at

the target region by co-transfection of Cas9 RNP and donor template. Screening PCR (Figure 4.19) and immunoblotting (Figure 4.21) indicated that monoclonal Hek-S₂₆₉₋₂₇₇-G₆(SGG)₃-1D3 cells potentially carry both the wild-type *B2M* allele and PLB-inserted *B2M* allele corresponding for B2M-like and PLB-like products, respectively. These proteins were also detected in the supernatant, suggesting they can form the complexes with HLA class I followed by transport to the cell surface and release during endocytic recycling of MHC-I molecules [355]. Both B2M-like and PLB-like proteins have enhanced expression of both HLA-B7, HLA-A2 and other MHC-I molecules. Immunoprecipitation experiments can be considered for future work to compare the interactions of B2M and PLB molecules with MHC-I complexes and remaining antigen processing complexes.

In contrast to Hek-S₂₆₉₋₂₇₇-G₆(SGG)₃-1D3 clone, Hek-S₂₆₉₋₂₇₇-G₆(SGG)₃-2B1 only expressed PLB-like molecules, suggesting one *B2M* allele were repaired following HDR pathway leading to PLB insertion, and the other repaired following NHEJ pathway which are potentially leading to the absence of wt *B2M* caused by gene knock-out (Figure 4.21). Although the Hek-S₂₆₉₋₂₇₇-G₆(SGG)₃-1D3 cell line has low surface expression levels of MHC-I, that of HLA-A2 was significantly higher than wt HEK cells. These observations suggest PLB molecules can selectively bind to peptide-specified HLA-I heavy chain only. Peptidome analysis can be considered for future work to investigate and compare the peptide spectrum between normal and gene-edited HEK cells. Furthermore, the function of PLB molecules should be evaluated via the recognition of peptide-specific CTL clones.

Stable expression of PLB molecules is a promising approach in loading specific antigen onto MHC molecules of APCs such as DC cells which can be used as cellular

therapies for infectious diseases, cancer or even autoimmune diseases. While conventional DCs (cDCs) present processed antigen by MHC-I to naïve CTLs and contribute to the maturation of CTLs, tolerogenic DC (tolDC) have recently been demonstrated to actively silence autoreactive T cells and reinstate immune tolerance [356, 357]. The MHC-I antigens expressed by both cDCs and tolDCs, have crucial roles in directing the action of CTLs. The current approaches for loading target antigen involve co-culturing peripheral blood-derived DCs with tumor lysates or synthetic peptides in combination with adjuvants before reinjecting into the same patient [358-360]. However, these above methods may face certain difficulties. In the case of tumor lysis, the peptides presented by DCs after exposing to tumor lysate may not always be tumour associated antigens or could be non-specific antigens which can affect the stimulation of CTLs. On the other hand, while the portion of empty MHC-I complexes, which normally reside in the ER in a peptide-receptive state by tapasin, are very low at the cell surface [54], synthetic peptides must compete with endogenous peptides to be loaded. Therefore, the PLB methodology can enhance the possibility of generating stable expression of the predesigned specific peptide and formation of MHC I class with antigens against pathogens and tumours for CTLs or autoantigens which help regulatory T cells induce antigen-specific tolerance.

Effective therapeutics requires the understanding of safety profiles of genome editing systems. One of the concerns that nuclease systems such as CRISPR/Cas9 bring, is its specificity. During the cellular repair process of DSBs, undesirable off-target editing can cause local mutations, genomic rearrangement or large deletions [361]. In this study, potential off-target sites were identified using computational prediction tools and surveyed by amplicon deep sequencing from polyclonal gene-edited cells. Off-target

analysis results indicated that indels were also not detected in two screened off-target loci of sgRNA.B2M.1.3 which were used for all HDR experiments for insertion of the PLB. Karyotyping or Chromosomal microarray analysis (CMA) can be considered for future work to detect large mutations, microdeletions or duplications. Furthermore, Whole Genome Sequencing (WGS) would be necessary to reveal all off-target mutations caused by CRISPR/Cas9 system at the stage of clinical trials and/or application.

In summary, this is the first study reported using the CRISPR/Cas9 system for the generation of PLB proteins which are stably expressed *in vivo* and provide pathogenic peptides to MHC-I complexes. My system, consisted of a Cas9 RNP and donor template, which precisely modifies the human *B2M* locus to successfully facilitate insertion of desired peptides and linker into the target region. I identified the introduction of inadvertent intronic motifs as an interruption factor of protein expression in large insertions of gene sequences which can be overcome by introducing silent mutations to remove these nucleotide motifs. For future work, the function of PLB needs to be evaluated by measuring CTL responses against these molecules. The stable expression of PLB protein holds promise for cellular therapy of cancer, infectious and autoimmune diseases.

CHAPTER 5

**MODIFICATION OF THE *HLA-B27* GENE BY
CRISPR TECHNOLOGY**

Chapter 5 Modification of the *HLA-B27* gene by CRISPR technology

5.1 Chapter aims & objectives

Several observations suggest that specific residues participate in the misfolding of HLA-B27 HC, whether it is the formation of HC-dimers or misfolding of the monomeric protein. Therefore, with the advent of gene editing, it is possible that making subtle changes to the HLA-B27 HC sequence could alleviate protein misfolding and possibly act as a therapeutic intervention strategy. Therefore, the above unanswered questions allowed me to think about changing the amino acid sequence of HLA-B*27:05 to create models for potential AS treatment. Therefore, my goal was to use the CRISPR/Cas9 gene-editing tool to introduce nucleotide changes to the HLA-B27 sequence. These changes included cysteine at p67 to serine (C67S) and aspartic acid at p116 to histidine (D116H). By targeting C67, it was anticipated that this could potentially reduce the ability of HLA-B27 to dimerise. As the 09 subtype is not associated with disease, supposedly due to the structural difference compared with 05 within the binding groove region and as peptide presentation was largely overlapping with 05 [362], I decided to also target D116. In addition, the disulfide bridge between Cys101 and Cys164 is important in the folding of MHC-I molecules, therefore, a C101S mutant will be generated using CRISPR/Cas as a control to assess general misfolding of HLA-B27 [130].

To achieve the aim, several objectives were established, including:

1. AS patients would only express 1-2 HLA-B27 alleles. Therefore, I wanted to generate cell lines expressing 1-2 copies of the mutated HLA-B27 alleles to assess their biochemical characteristics as a reference to HLA-B27

molecules that were targeted by CRISPR. I wanted to generate pFRT/B27:05 plasmids containing C67S, C101S and D116H which could be integrated into a HeLa cell line that would express 1-2 copies of the mutated molecules using the Invitrogen Flip-in system.

2. Targeting MHC alleles with a view of changing single amino acids has yet to be achieved using CRISPR/Cas. Therefore, I needed to design and optimize CRISPR/Cas9 systems targeting the endogenous HLA-B*27:05 allele at p67, 101 and 116.
3. To assess the expression and biochemical impact of the above edited HLA-B27 molecules, I needed to generate cell lines stably expressing B27 molecules containing the different variants generated by CRISPR/Cas9 gene editing. The cell lines would then be subjected to flow cytometric analysis to determine cell surface expression and immunoblotting to examine any potential biochemical changes.

5.2 Results

5.2.1 Generation of HLA-B27 mutants C67S, D116H and C101S by Site Directed Mutagenesis

One of the strategies I want to use for therapeutic interventions was to quickly assess successful editing. Therefore, Restriction Fragment Length Polymorphism (RFLP) was employed as part of mutagenesis approach to ensure the choice of changes (C67S, D116H and C101S) could be employed in the gene editing approach without any adverse effects on HLA-B27. Besides introducing the amino acid substitution at pC67, C101 and D116, I inserted two or three silent mutations into the HLA-B*27:05 sequence for each position. Silent mutations were expected to introduce a restriction enzyme site without affecting the amino acid residue. Using web based Watcut silent mutation analysis (<http://watcut.uwaterloo.ca/>), I found the appropriate nucleotide substitutions which could introduce cleavage sites for the following restriction enzymes EcoRV, BamHI and NsiI for C67S, C101S and D116H, respectively (Figure 5.1).

Amino acid change	Primer-Template Duplex 1	Primer-Template Duplex 2
67.C>S	<p style="text-align: center;">C</p> <p>acgggagacacagatctgcaaggccaaggcacag</p> <p> </p> <p>3'-cctctgtgtctagtcgttccgggtccgt-5'</p> <p style="text-align: center;">S</p> <p>5'-gggagacacagatcagcaaggccaaggca-3'</p> <p> </p> <p>tgccctctgtgtctagaaggtccgggtccgtgtc</p>	<p style="text-align: center;">I C</p> <p>ggagtattgggaccgggagacacagatctgcaaggccaaggcacaga</p> <p> </p> <p>3'-cataaccctggccctctgtgtctatagattccgggtccgtg-5'</p> <p style="text-align: center;">I S</p> <p>5'-gtattgggaccgggagacacagatctctaaggccaaggcac-3'</p> <p> </p> <p>gctcataaccctggccctctgtgtctagaaggtccgggtccgtgtc</p> <p style="text-align: center;">EcoRV</p>
101.C>S	<p style="text-align: center;">C</p> <p>caacctccagaatatgtatggctgcgacgtggggccgg</p> <p> </p> <p>3'-ggaggctcttatacataccgaggtgcacccccg-5'</p> <p style="text-align: center;">S</p> <p>5'-cctccagaatatgtatggctccgacgtggggc-3'</p> <p> </p> <p>gtgggaggtcttatacataccgacgtgcacccccggcc</p>	<p style="text-align: center;">Y G C</p> <p>gtctcacaccctccagaatatgtatggctgcgacgtggggccgga</p> <p> </p> <p>3'-agtgtgggaggtcttatacatgcttagctgcacccccgg-5'</p> <p style="text-align: center;">Y G S</p> <p>5'-tcacaccctccagaatatgtacggatccgacgtggggcc-3'</p> <p> </p> <p>gagagtgtgggaggtcttatacataccgacgtgcacccccggcc</p> <p style="text-align: center;">BamHI</p>
116.D>H	<p style="text-align: center;">D</p> <p>cgcggtaccaccaggacgcctacgaaggca</p> <p> </p> <p>3'-cccatggtggtcgtgaggatgctgc-5'</p> <p style="text-align: center;">H</p> <p>5'-gggtaccaccagcagcctacgacg-3'</p> <p> </p> <p>gcgccatggtggtcctgaggatgctgcggt</p>	<p style="text-align: center;">D A</p> <p>cccggtaccaccaggacgcctacgaaggcat</p> <p> </p> <p>3'-cgccatggtggtcgtgacgtatgctgcccgttcct-5'</p> <p style="text-align: center;">H A</p> <p>5'-gcggtaccaccagcatgcatacgaaggca-3'</p> <p> </p> <p>agggcccatggtggtcctgaggatgctgcccgttcctaat</p> <p style="text-align: center;">NsiI</p>

Figure 5.1. Nucleotide substitutions introducing the desired amino acid changes and restriction enzyme sites within the HLA-B*27:05 sequence.

The HLA-B*27:05 nucleotide sequences are highlighted in green. Desired variants were introduced into HLA-B*27:05 by Quikchange SDM methodology using two sets of complementary overlapping primers (highlighted in yellow): Primer-Template Duplex 1 for amino acid substitutions and Primer-Template Duplex 2 for both amino acid substitutions and silent mutations to create restriction enzyme sites for RFLP analysis. Combination of nucleotide substitutions introduced by Primer-Template Duplex 2 can generate a new restriction enzyme site into the HLA-B*27:05 sequence.

The aim of this section was to generate cell lines containing different HLA-B*27:05 variants using the Invitrogen Flip-in technology (<https://www.thermofisher.com/>). I was provided with the Flip-In-HeLa cells (HeLa-H2Z) expressing 2 Flp Recombination Target (FRT) sites [363] and the pcDNATM5/B27/FRT/V5 (Figure 5.2A) which was previously generated by Dr Darren Nesbeth, UCL. The FRT site sequence (5'-GAAGTTCCTATTCTctagaaaGtATAGGAACTTC-3') allows integration and expression of the gene of interest (GOI) in mammalian cells. PcDNATM5/B27/FRT/V5 plasmid was constructed from a pcDNATM5/FRT expression vector (Appendix 3, supplementary data 1B) with a carboxy-terminally V5-tagged HLA-B27. Expression of HLA-B27 was controlled by the human CMV promoter. pcDNATM5/B27/FRT/V5 plasmid also contains the hygromycin B resistance gene and a FRT site. V5 is a short peptide tag which was fused to the C' terminus of HLA-B27 and used for detection of this heavy chain protein. I then introduced nucleotide substitutions into pcDNATM5/B27/FRT/V5 using Quik-change Site-Directed Mutagenesis (SDM) methodology.

Prior to performing SDM, primer pairs F4-B27 and R4'-B27, were designed to amplify the entire HLA-B27 cDNA sequence from the pcDNATM5/B27/FRT/V5 plasmid. PCR using F4-B27 and R4'-B27 primers was optimized by gradient PCR, which amplified a

single amplicon (913 bp) of the desired size at annealing temperatures between 60°C and 62.5°C (Figure 5.2B). Therefore, F4-B27 and R4'-B27 primers were used for later experiments with pcDNA™5/B27/FRT/V5 with annealing temperature of 61°C.

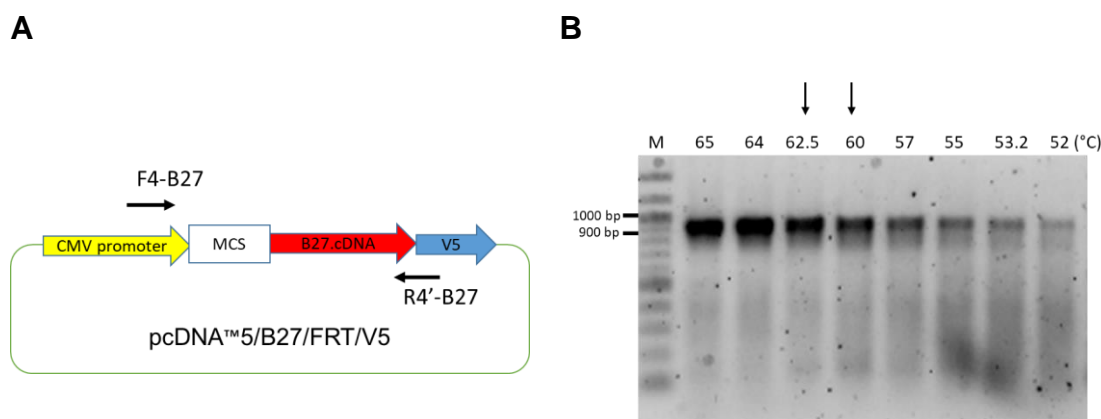


Figure 5.2. Specific primers for amplifying cDNA of B27 from the pcDNA™5/B27/FRT/V5 plasmid.

(A) Graphic illustration of pcDNA™5/B27/FRT/V5 plasmid and primer pairs amplifying B27 cDNA. F4-B27 and R4'-B27 primers amplified from nucleotide position 549 within the CMV region and nucleotide position 21 within the V5 sequence. The amplicon has a length of 914 bp and covers the whole cDNA sequence of B27. (B) Agarose gel electrophoresis image for the gradient PCR of B27 cDNA. B27 cDNA were amplified using F4-B27 and R4'-B27 primers at different annealing temperatures. PCR bands at 60°C and 62.5°C were observed to be the two lowest annealing temperatures providing an amplicon with correct size (913 bp) and without non-specific bands. Data are presented of one experiment.

The B27 plasmid with different variants were generated using the Quikchange site-directed mutagenesis protocol. Two clones of each mutated plasmid were isolated and expanded for DNA extraction. The pcDNA™5/B27/FRT/V5 plasmids containing desired variants were sequenced using the F4-B27 primer by Sanger sequencing (Sequencing Service Dundee University). Sequencing results showed that pcDNA™5/B27/FRT/V5 had no additional nucleotide substitutions and that HLA-

B*27:05 cDNA sequence contained the desired single- or multiple- substitutions (Figure 5.3A). Plasmids with single-nucleotide substitutions were named pcDNA™5/B27/FRT/V5-C67S, pcDNA™5/B27/FRT/V5-C101S and pcDNA™5/B27/FRT/V5-D116H. Plasmids with multiple-nucleotide substitutions were provided with the same name and asterix pcDNA™5/B27/FRT/V5-C67S*,

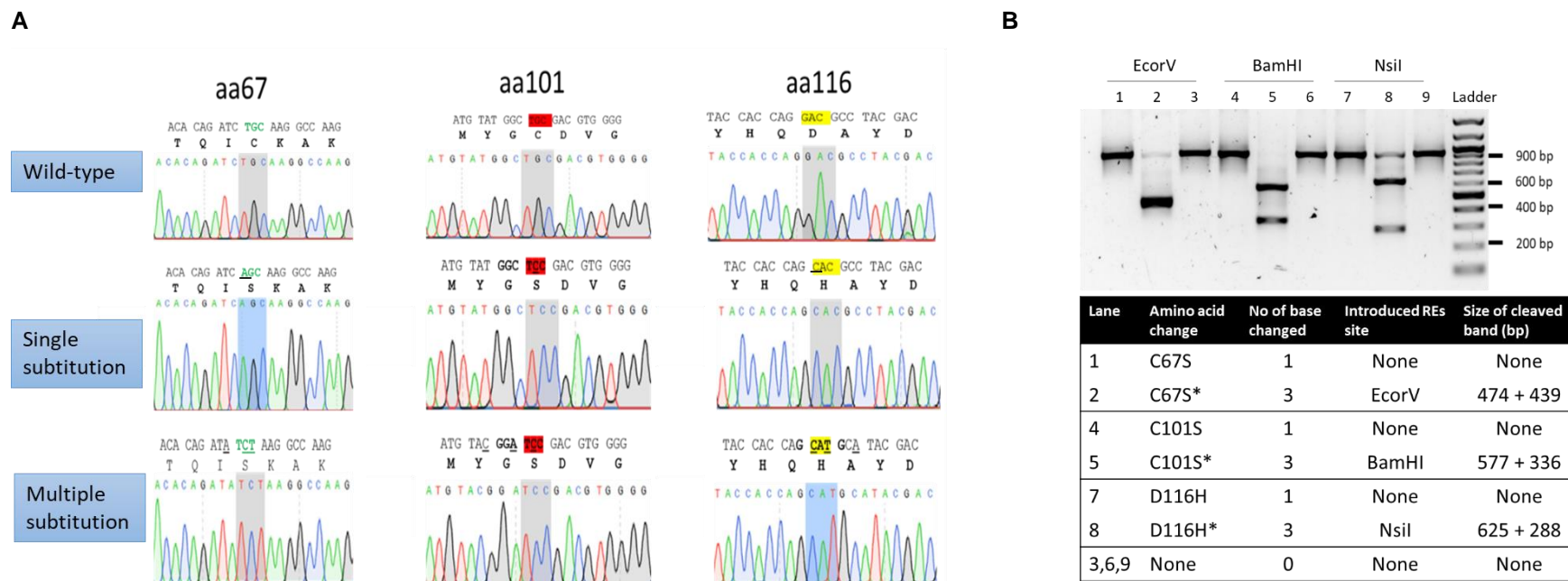


Figure 5.3. Site-directed mutagenesis of pcDNA™5/B27/FRT/V5 introducing C67S, D116H and C101S changes.

(A) Sanger sequencing of plasmids containing HLA-B*27:05 wild-type, single-substitution and multiple-substitution sequences. Single substitution sequences have only one nucleotide change (underline) and multiple-substitution sequences have three nucleotide changes. Triplet codons and corresponding amino acids at position 67, 101 and 116 were highlighted in green, red and yellow, respectively. (B) RE digests of PCR amplicons of HLA-B*27:05 wild-type and mutated plasmids resolved by DNA 1.2% (w/v) agarose gel electrophoresis. PCR amplicons of 913bp were generated by the primer pair of F4-B27 and R4-B27, treated with corresponding REs: EcoRV for C67S*, BamHI for C101S* and Nsil for D116H*. Undigested PCR products of pcDNA™5/B27/FRT/V5 was used as a negative control (lane 3,6 and 9). Data are presented of one experiment.

pcDNATM5/B27/FRT/V5-C101S* and pcDNATM5/B27/FRT/V5-D116H*. To ensure the introduced silent mutations can generate cleavage sites for the desired restriction enzyme sites, B27 cDNA amplicons were generated from wild-type, single-substitution and multiple-substitution plasmid templates and digested with the corresponding RE (EcoRV for C67S*, BamHI for C101S* and NsiI for D116H*) (Figure 5.3B). No cleavage products were detected for wild type B27 (lane 3, 6 and 9) and single-substitution amplicons (lane 1, 4 and 7). Restriction enzyme digestion of B27 multiple-substitution amplicons did generate cleavage products (lane 2, 5 and 8) indicating that RFLP analysis could successfully detect introduced mutations.

To generate stable cell lines expressing HLA-B*27:05, I co-transfected pcDNATM5/FRT constructs containing the desired HLA-B27 mutation and pOG44 plasmid, which encodes the Flippase protein, into the HeLa-H2Z cell lines. Transfected cells were grown in selection medium containing both hygromycin and zeocin to ensure that I could select cell lines not only with 2 copies but also single copies of the desired mutation. To check the success of the Flp-in transfection, equal number of each cell line (5×10^5) were harvested and lysed in 1% NP-40 lysis buffer. After resolving reduced lysates by SDS-PAGE, samples were immunoblotted with the anti-V5 pK and GAPDH antibodies (Figure 5.4). Immunoblotting with the house-keeping protein GAPDH (~37 kDa) revealed that there was some unequal protein loading (lanes 5, 7 and 9). Immunoblotting with the anti-V5 antibody was expected to detect the fused protein of HLA-B*27:05 heavy chain (~40.5 kDa) and V5 C' terminal tag (~1.4 kDa) with the band size of ~41.9 kDa. Immunoblotting revealed expression of all mutants indicating successful integration into FRT sites of the HeLa-H2Z host cell.

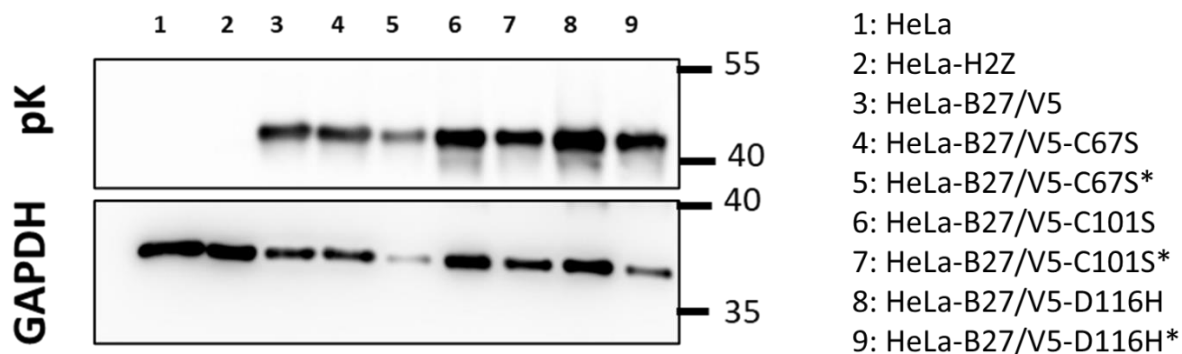


Figure 5.4. Confirmation of mutated-HLA-B27 integration into HeLa-H2Z.

HeLa cells, untransfected HeLa-H2Z, HeLa-B27/V5 and HeLa-B27/V5 containing mutations (C67S, C67S*, C101S, C101S*, D116H and D116H*) were analyzed for expression. Cell lines were lysed in 1% NP-40 and reduced samples were resolved by two separate 8% SDS-PAGE gels. One gel was immunoblotted with anti-V5 antibody pK for detection of HLA-B27/V5 tagged protein (~41.9 kDa). The second gel was immunoblotted with an antibody against the housekeeping protein, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) for reference. Data are presented of one experiment.

5.2.2 Investigating C67S, D116H and C101S.B27 expressing cell lines

After generating the HeLa cell lines carrying HLA-B27 and the C67S, D116H and C101S mutants, cell surface expression of HLA-B27 were determined by flow cytometric analysis using equal number of cells (2×10^5), stained with the ME1 antibody which binds specifically to the MHC-I complexes of HLA-B27, B22 and B7 heavy chain [334]. Beside HeLa and HeLa-H2Z, I also used the HeLa-E84 cell line, which contains an 'empty' pcDNA™5/FRT vector, as an additional negative control. Flow cytometry results (Figure 5.5A) revealed that as expected there was no HLA-B27 detected on HeLa, HeLa-H2Z and HeLaE84 cell lines. HLA-B27 was observed on HeLa-HLA-B27, D116H and C67S, but no expression of C101S. While cell surface expression of HLA-B27 was not detected in both C101S and C101* cells, the differences in cell surface

expression between each pair of C67S-C67S*, and D116H-D116H* were not significant. It suggests additional silent nucleotide mutations do not affect the cell surface expression of HLA-B27.

Although HLA-B27-D116H was detected at the cell surface, the levels of D116H (MFI: 34040) and D116* (MFI: 20885) was less than that detected for wild type (wt) HLA-B27 (MFI: 43,638) (Figure 5.5B). The above results indicate that the HLA-B*27:09 subtype has lower expression than HLA-B*27:05 subtype. Flow cytometry analysis revealed that fully assembled ME-1-reactive HLA-B27-C67S was also expressed at lower levels than the wt HLA-B27. While expression of B27.C67S (MFI: 9700) and B27.C67S* (MFI: 15292) were significantly lower than that of HLA-B27 (MFI: 43,638), both B27.C101 and B27.C101* were not observed to express HLA-B27 at the cell surface as expected. However, the C67S mutant was also detected at reduced levels compared to WT HLA-B*27:05 suggesting that C67 may have an impact on the folding and cell surface expression levels of HLA-B27.

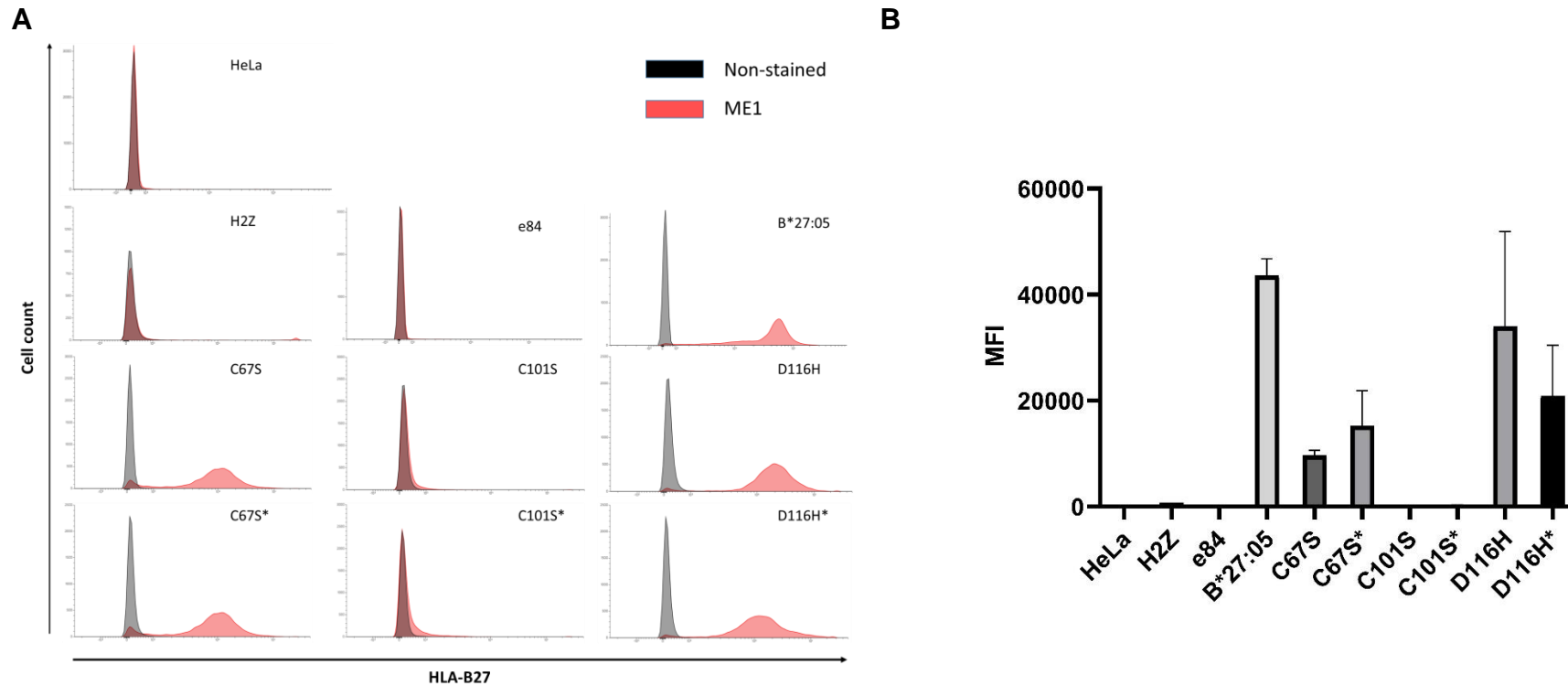


Figure 5.5. Flow-cytometry analysis of HeLa cell lines expressing HLA-B27 and B27 mutants.

(A) The flow cytometry histograms of surface HLA-B27 expression on HeLa, HeLa-H2Z, HeLa-E84, HeLa-HLA-B*27:05 and HeLa-B*27:05 mutants (C67S, C67S*, C101S, C101S*, D116H and D116H*). Surface HLA-B27 expression was analyzed by flow cytometry using ME1 antibody staining (red) and compared to the signal of non-stained samples (black). (B) The bar charts indicate the Mean Fluorescence Intensities (MFI) of HeLa-B*27:05 expression from two different experiments for HeLa-HLA-B*27:05 and HeLa-B*27:05 with C67S, C67S*, C101S, C101S*, D116H and D116H* mutations. The means of MFI were calculated from duplicated data.

In order to determine the dimerization of the various HLA-B27 molecules expressed by HeLa cells in this controlled expression system, equal number of cells (10^6) were pre-treated with N-ethylmaleimide (NEM) followed by lysis in 1% NP-40 detergent. NEM is a small organic compound containing the alkene that forms stable, covalent thioether bonds with free sulfhydryl groups such as those of reduced cysteines and prevents them from reforming disulphide bonds. Therefore, NEM pre-treatment can block the further formation of heavy chain dimers through disulphide binding between the thiol groups of two free cysteines. Non-reduced and reduced lysates were resolved on 10% SDS-PAGE followed by immunoblotting with HC-10 antibody, which can recognize and bind to unfolded HLA class I B and C alleles [364] and pK antibody, which can recognize and bind to the V5 tag. A further blot containing reduced lysates were probed for GAPDH as a loading control.

Immunoblotting with GAPDH (Figure 5.6A) indicated that there was approximately equal loading of protein between the cell lysates. Immunoblotting with HC10 on reduced cell lysates (Figure 5.6A) revealed that there was one isoform (~40 kDa) of MHC-I detected in HeLa and HeLa-E84, and two isoforms (~40 kDa and 42 kDa) of MHC-I by HeLa cells expressing HLA-B*27:05 wild type and cells expressing the C67S, D116H and C101S mutants. The lower Mw HC10 reactive band correlated with the endogenous HLA B/C alleles whilst the higher Mw isoform were possibly the HLA-B27 V5 tagged molecules. Immunoblotting of reduced cell lysates with pK antibody (Figure 5.6A and Figure 5.6B, left panel) confirmed that the HLA-I isoform with higher Mw were HLA-B27 and the respective mutants. Immunoblotting non-reduced cell lysates with pK revealed different conformers of HLA-B27 heavy chains. These

conformers were observed at different sizes which can be grouped into monomers with sizes of approximately 40 kDa and homodimers between 80-130 kDa.

Immunoblotting with HC10 and pK of non-reduced cell lysates of HeLa-B27 expressing cells (Figure 5.6 A, B-right image and C-right image) revealed three different high molecular weight bands (arrow i-iii) and three different low molecular weight bands (arrow iv-vi). While the low molecular weight proteins are representative of monomer conformers, the higher Mw are possibly different homodimers or heterodimer of HLA-B27 molecules and other proteins. The same number of bands were observed by HeLa-B27-D116H and -D116H* expressing cells. These observations suggest that HLA-B*27:09 does not exhibit significant changes with respect to HLA-B27 heavy chain dimer formation. However, different HC-dimer bands could be detected by cells expressing C67S and C101S. Only the highest Mw conformer (i), could be detected in cells expressing C67S, C67S*, C101S and C101S*, indicating that substitution of cysteine at p67 and p101 can alter the HC-dimer conformers detected. Additionally, immunoblotting with HC10 and pK of non-reduced cell lysates of HeLa-B27-C101S and C101S* (Figure 5.6 B-right image and C-right image) revealed that substitution of cysteine at p101 also led to the absence of one monomer conformation (isoform v). C101 has been described to play an important role in the folding of HLA-B27 heavy chain by participating in disulphide bond formation with C164 [130]. Without the C101S-C164 disulphide bond, HLA-B27 does not fold appropriately and is not capable of loading the peptide and presenting to the cell surface. Therefore, it suggests that only isoform v can be the correct intramolecular folded molecules, while other isoforms including both homodimer and monomer (isoform iv and vi) were aberrant or represent incompletely folded molecules.

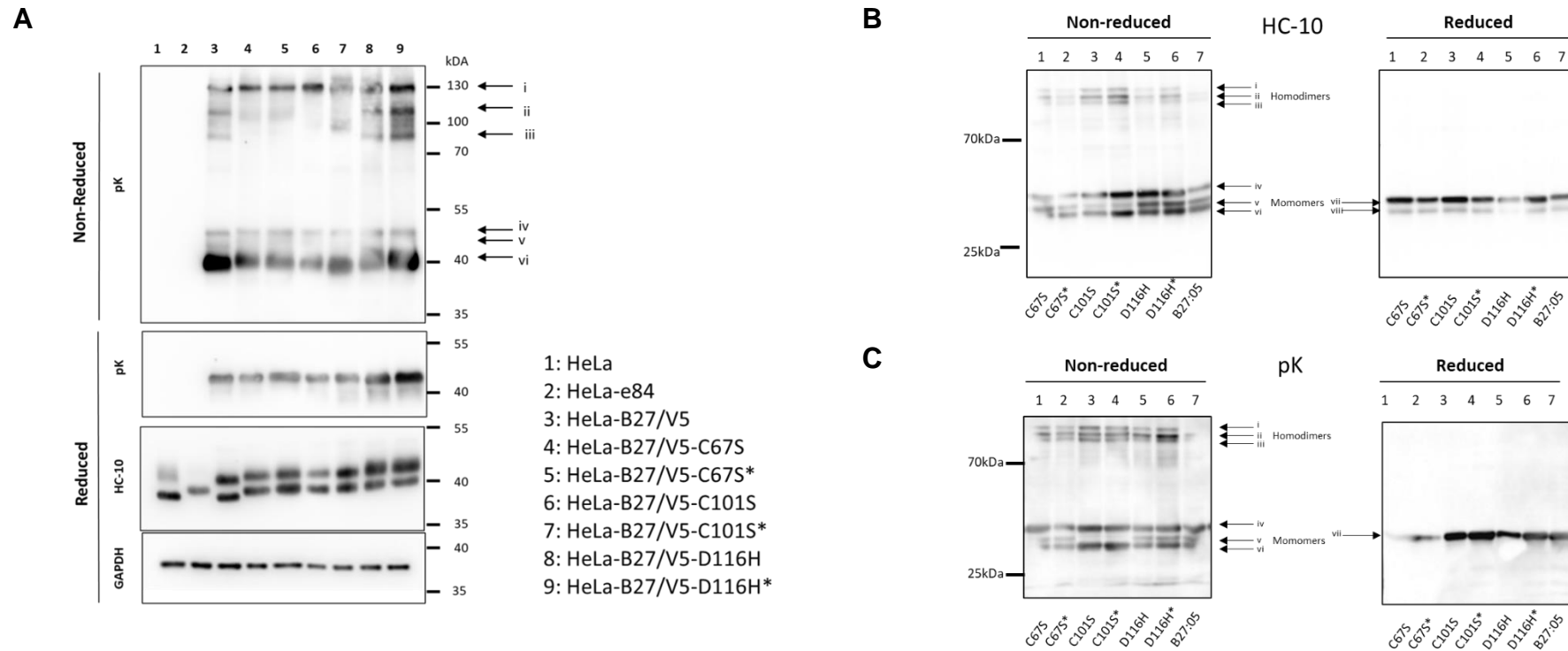


Figure 5.6. Immunoblotting analysis of MHC-I expression on HeLa cell lines.

HeLa, HeLa-H2Z, HeLa-E84, HeLa-HLA-B*27:05 and HeLa-B*27:05 mutants (C67S, C67S*, C101S, C101S*, D116H and D116H*) were pretreated in NEM (20mM) and lysed in 1% NP-40 detergent. (A) Non-reduced cell lysates were resolved on a 10% SDS-PAGE gel followed by immunoblotting with anti-V5 antibody pK. Reduced cell lysates were resolved on 10% SDS-PAGE gels followed by immunoblotting with pK and HC10 which detects partially folded/misfolded B/C alleles, and with anti-GAPDH antibody. (B, C) Non-reduced and reduced cell lysates were resolved on 10% SDS-PAGE gel followed by immunoblotting with HC10 (B) and pK antibody (C). Arrow i-viii indicate different forms of MHC-I molecules. Data are presented of one experiment.

In summary, I have generated HeLa cell lines containing HLA-B27 alleles with different mutations using site directed mutagenesis and the Flip-In system. The established cell lines were used for detecting biochemical differences which can be used as a reference for gene editing.

5.2.3 CRISPR/Cas9 system targeting HLA-B*27:05

The next aim was to directly introduce the nucleotide substitutions into HLA-B*27:05 using CRISPR/Cas9 for generating cell lines expressing the above different variants of HLA-B27. To facilitate HDR on the genomic HLA-B*27:05 sequence, the target sites need to be cleaved by Cas9 RNP to generate single or double stranded breaks followed by DNA repair using a donor DNA template. CRISPR-mediated gene editing requires DSBs generated by the cleavage by an RNP complex of Cas9 protein and sgRNAs. The high polymorphism of the HLA class I locus presents certain challenges especially with specific targeting and off-target effects. Therefore, different sgRNAs were designed to target different sites of HLA-B*27:05 and optimized by initial performance on cell lines expressing HLA-B27.HC cDNA, prior to moving onto genomic DNA sequences.

As the most widely used method for genome manipulation, sgRNA-guided *S. pyogenes* Cas9 (SpCas9) system was harnessed to target the HLA-B27 cDNA sequence. The design of sgRNAs used the following criteria appropriate for employing *spCas9*: (i) the length of 20 nucleotides; (ii) DSBs located close to the desired target sequence (<20nt); (iii) high score of off-target activity (>60) and (iv) on-target activity (>40). Based on the *spCas9*'s PAM motif (NGG), appropriate crRNA can be identified within the target regions. While the score of off-target activity predicts the inverse probability of Cas9 off-target binding, the on-target score indicates the cleavage efficiency of Cas9 RNP. Therefore, higher scores of on- and off-target means the sgRNA has a reduced chance of binding to non-specific sequences and a higher efficiency to generate DNA breaks at the desired target sequence [365]. Using Benchling bioinformatics software, five sgRNAs were chosen to target HLA-B27 at the

amino acid position C67 (sg272F and sg278F), C101 (sg359F and g379F), and D116 (sg426F) (Figure 5.7).

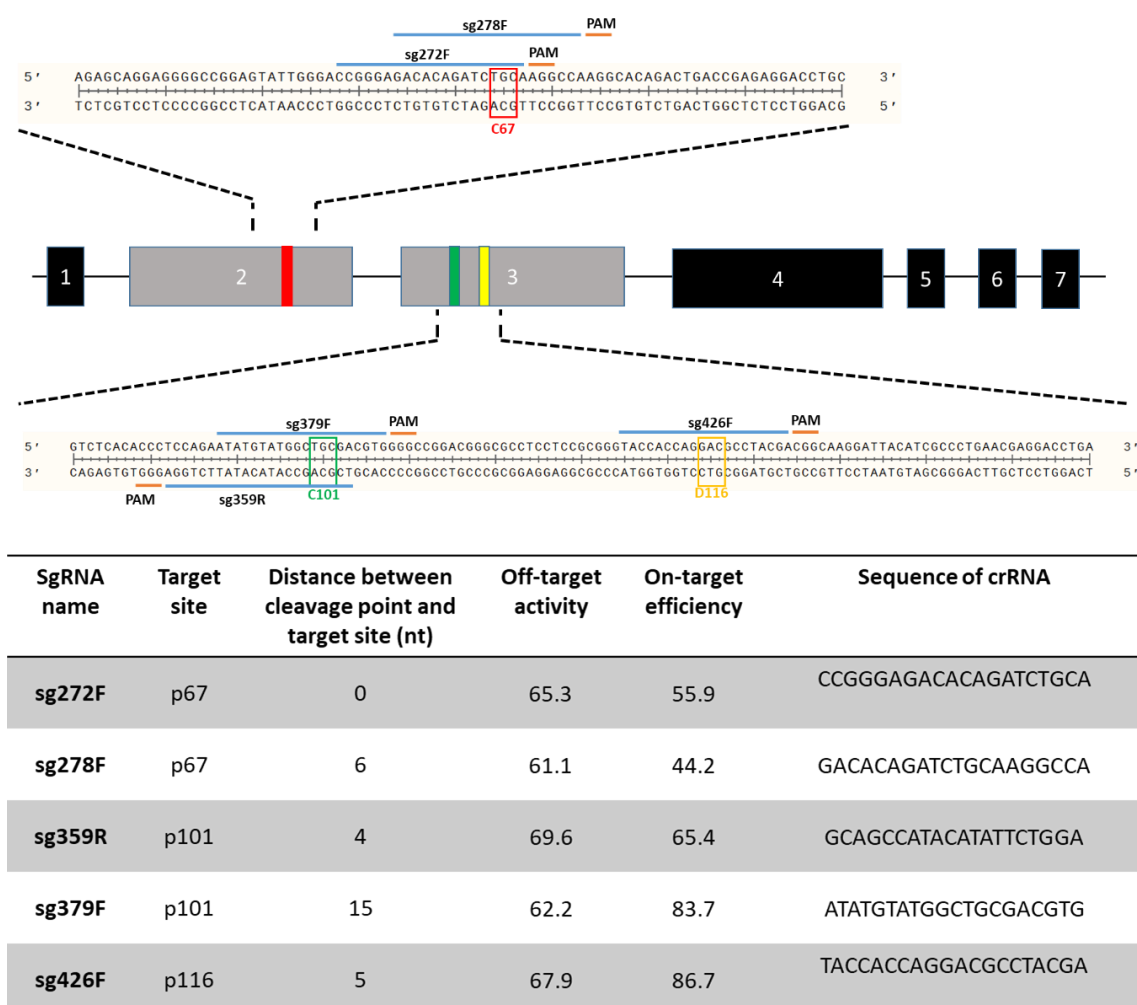


Figure 5.7. Selection of sgRNAs targeting the HLA-B*27:05 allele.

Genomic HLA-B*27:05 contain 7 exons in which C67 is located within exon 2 and C101 and D116 are located within exon 3. Five crRNAs (blue line) were designed to target the HLA-B27 sequence at amino acid C67 (red), C101 (green) and D116 (yellow). All 20-nt crRNAs were designed so that they were directly upstream of a protospacer adjacent motif (PAM) site (NGG). The table stated the details of each sgRNAs.

In order to evaluate the function of the sgRNAs, three primer pairs were designed to amplify different regions within the expression cassette of HLA-B*27:05 cDNA (Figure 5.8A). At least one primer of each pair was in the CMV promoter or V5 tag to avoid the amplification of other HLA class I alleles. While the amplicon of primer pair 1 was expected to cover sequences encoding C101 and D116 of HLA-B27, that of primer pair 3 were expected to cover C67. On the other hand, the amplicon of primer pair 1 were expected to cover all target sites.

The function of the designed sgRNA was evaluated via the IVT cleavage assay (Figure 5.8B). In this assay, naked DNA amplicons of the target sites were treated with the RNP complex of Cas9 nuclease and sgRNAs and separated by agarose electrophoresis. If Cas9 RNP can recognize the target site via the interaction between crRNA and target sequence, Cas9 protein will induce the DSBs and cleave the PCR amplicons into 2 separate bands which can be visualized by agarose electrophoresis. HPRT sgRNA (IDT) was used as a positive control to validate reagents or troubleshoot the experiments using crRNA:tracrRNA duplexes. Sample 1 of the IVT assay indicated that the RNP complex of HPRT + sgRNA cut the HPRT amplicon (1083 bp) to two smaller bands (827 and 256 bp). A negative control gRNA (IDT) was used containing a 20 nt crRNA that was computationally designed to be non-targeting to the human reference genome. Sample 2, 8 and 13 of the IVT assay showed the presence of PCR amplicons only but no cleaved bands. Cleavage products were observed with all samples treated with sgRNAs targeting HLA-B27 in lanes 5, 6, 8, 9 and 10 (Figure 5.11B). The repeated IVT assay (Figure 5.8C) confirmed that the designed CRISPR guide RNAs were functional and ready for later experiments.

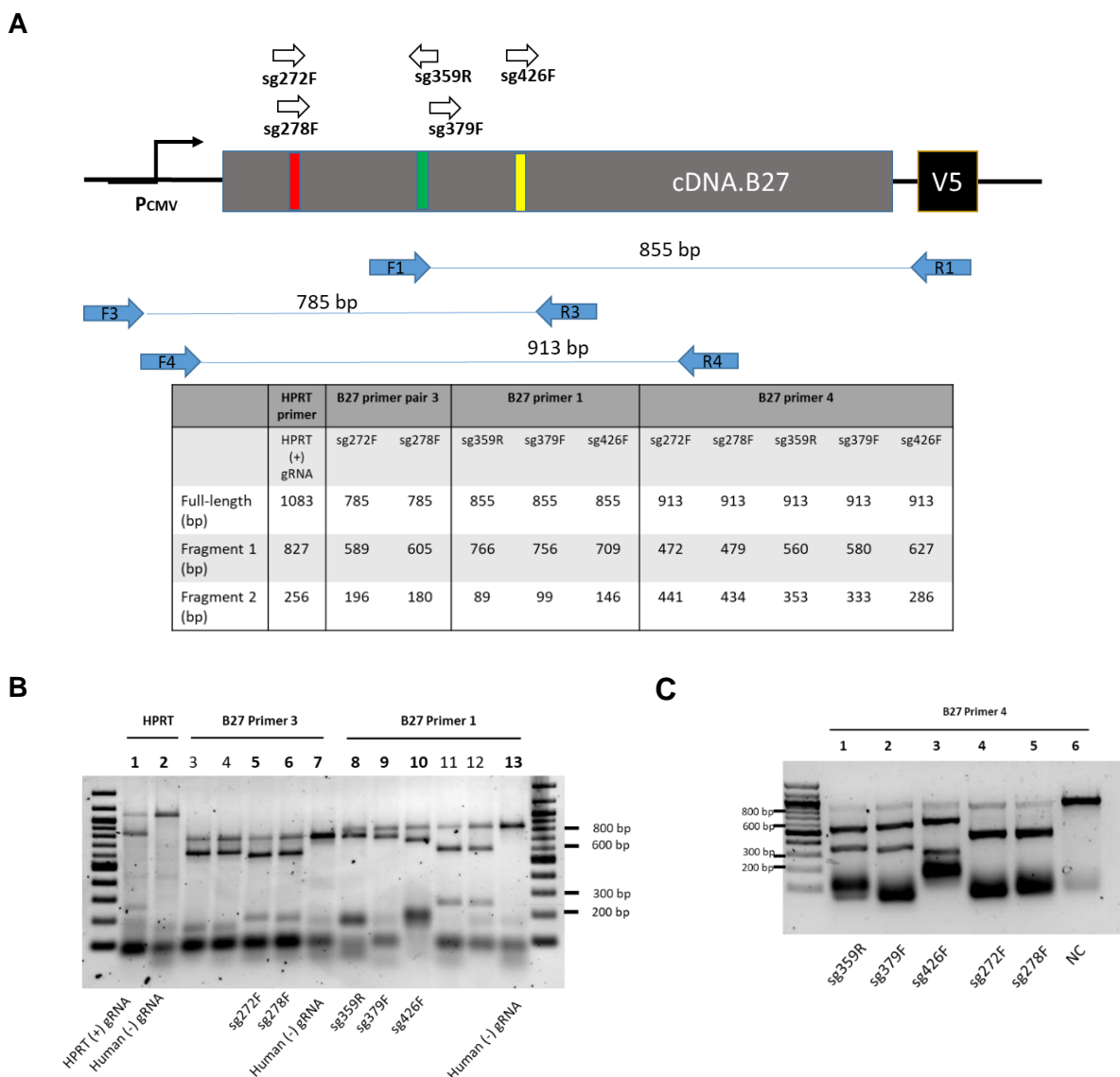


Figure 5.8. *In vitro* cleavage assay of Cas9 RNP targeting HLA-B27 cDNA

(A) The cassette of HLA-B27 expression was composed of CMV promoter, HLA-B27 cDNA and V5 tag sequences. Five sgRNAs (white arrows) targeted B27 at C67, C101 and D116. Three primer pairs (blue arrow) were designed for amplification of HLA-B27. The bottom table stated the size of PCR amplicons of each primer pairs and expected cleavage bands by Cas9 RNPs. (B, C) 1.2% (w/v) agarose electrophoresis image indicated *in vitro* cleavage results of Cas9 RNPs. HPRT positive and negative gRNA were used as positive and negative controls for the assays. Data are presented of one experiment.

For the next step, I evaluated the ability of Cas9 RNPs to target *in vivo* the HLA-B27 sequence using an *in vivo* (IVV) cleavage assay (Figure 5.9). RNPs of Cas9 proteins and sgRNAs were delivered into HeLa-B27 cells by lipofection. One day post transfection, genomic DNA was extracted and used as a template to produce the PCR amplicon using HLA-B27 specific primers. The PCR amplicons were denatured and reannealed to allow heteroduplex formation between mutated and wt DNA sequences followed by incubation with T7 endonuclease I. CRISPR-mediated indel mutations were then detected based on the T7EI-treated amplicons. Cleavage products were detected in samples transfected with RNP of sg359R (lane 1), sg426F (lane 3), and sg278 (lane 9). The faded band in lanes 2 and 8 indicated that the IVV cleavage efficiency of sg379F and sg272F were lower compared to other sgRNAs. These results indicated that the designed sgRNAs including sg278F, sg359R and sg426F could be used to introduce DSB which could facilitate DNA repair following HDR for gene-insertion within the HLA-B27 gene at positions C67, C101 and D116, respectively.

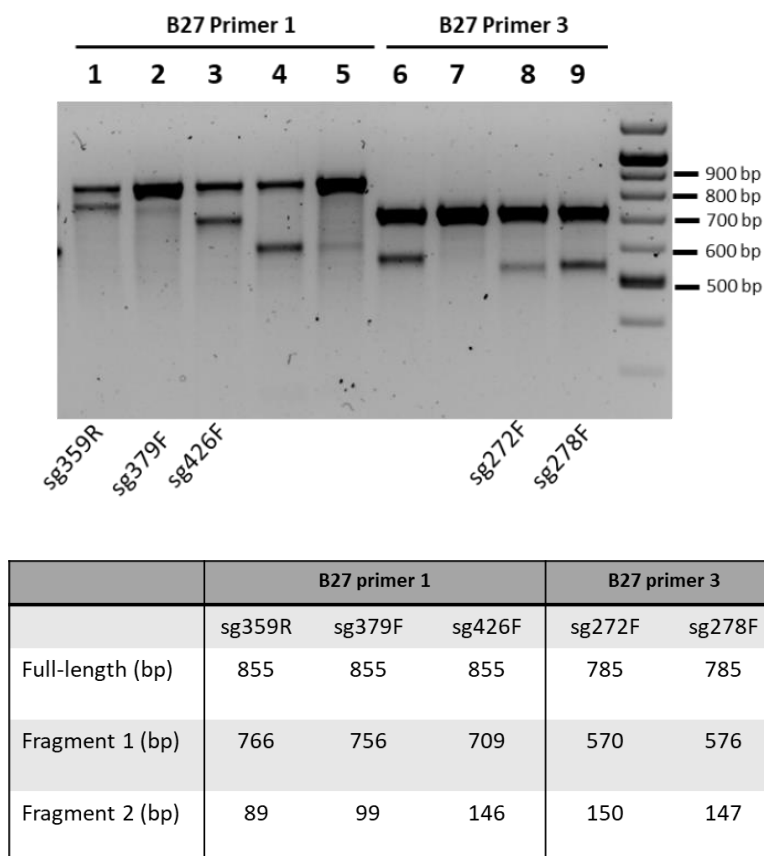


Figure 5.9. *In vivo* T7EI cleavage result of Cas9 RNPs targeting HLA-B27.

Agarose electrophoresis results of T7EI digestion of PCR amplicons generated from HeLa-B27 cells transfected with Cas9 RNPs targeting HLA-B27. After thermal cycling (10 min. 95°C; 95–85°C (ramp rate: –2°C/sec); 85–25°C (ramp rate: –0.3°C/sec), amplicons with indel mutations formed the heteroduplex and were digested by T7EI into smaller bands. Data are presented of one experiment.

5.2.4 Selection of cell lines for HDR

CRISPR-mediated insertion of a desired sequence requires cells to repair DSBs using endogenous HDR mechanisms but most cells exhibit a preference for the NHEJ pathway, which is more robust, predominant and flexible compared to HDR [366]. Therefore, it is necessary to identify appropriate cell types for HDR purposes before applying the CRISPR/Cas9 systems to introduce desired mutations within the HLA-

B27 cDNA. *B2M* locus was chosen as the target for this optimization experiment due to the highly conserved of this gene and also relate to MHC-I complex. Different cell lines including HEK, HeLa and K562 were transfected with a Cas9 RNP targeting exon 1 *B2M* locus and donor template (used in Chapters 3 and 4) (Figure 5.10). Optimized SgRNA.B2M.1.3 was then used to form the RNP with the Cas9 nuclease. The 135-nt donor template (ssODN.B2M.SilentM) contains two 65 bp homology arms and a 7 bp insertion which had 3 nucleotides substituted to introduce 3 silent amino acid mutations into the *B2M* allele (Figure 5.10). These changes to the original sequence introduce restriction enzyme sites (*Xba*I) allowing for RFLP analysis to evaluate whether HDR had been achieved.

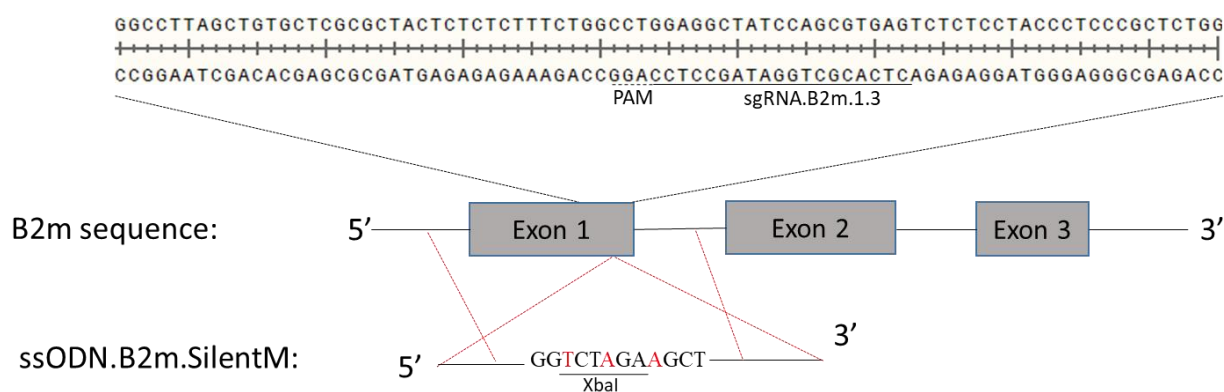


Figure 5.10. Design of sgRNA and donor template for HDR targeting exon 1 of the *B2M* locus.

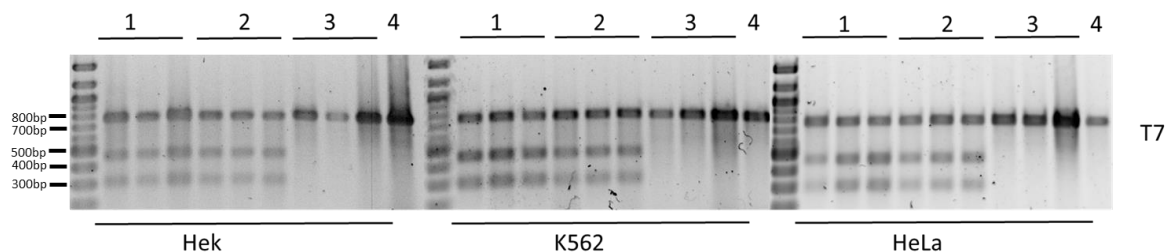
Genomic *B2M* contains 3 exons located on the long arm of chromosome 15 at 15q21.1. sgRNA.B2M.1.3 were designed to target the end of exon 1 (Chr15:25003800). Donor templates (ssODN.B2M.SilentM) were used in combination with Cas9 RNPs to introduce HDR-mediated insertion at the target sequence. The 7 nt insertion sequence contains silent mutations and a *Xba*I RE site. The homology arms were 75 nt in length.

The HDR within the B2M locus was facilitated by co-transfecting the dsODN donor templates and RNP of Cas9 protein and sgRNA.B2M.1.3 using electroporation with the Amaxa Nucleofector system (Lonza). Equal numbers (10^6) of HeLa, HEK and K562 cells were transfected with ssODN and RNP complex at a molar ratio of 4:1 followed by the incubation in complete culture media with added HDR Enhancer (group 1). Cells transfected with Cas9 RNP only (group 2) and donor template only (group 3) were used as negative controls for HDR and DSBs, respectively. All transfections were performed in triplicate on Hek, K562 and HeLa cell line. Genomic DNAs were extracted from one day post-transfected cells and used for preparing amplicons for the T7EI assay using the F1.B2M.exon1 and R1.B2M.exon1 primers. PCR products of DNA extracted from wild-type cells (group 4) were used as a positive control for the PCR reaction. The T7EI assay (Figure 5.11B) showed that bands of 458 and 304 bp resulting from T7EI cleavage were detected in HeLa, K562 and HEK samples transfected with Cas9/sgRNA.B2M.1.3 RNP (group 1 and 2). These data indicated that Cas9/sgRNA.B2M.1.3 RNP complexes can introduce DSBs in all cells and lead to indel mutations (without donor template) or desired insertion (with donor template). The RFLP assay (Figure 5.11C) though, showed cleaved products with the sizes of 455 bp and 307 bp were detected only within K562 cells transfected with RNP and donor template (K562_group 1). These results indicates that the K562 cells preferentially repair DSBs following the HDR pathway.

A

Sample	1	2	3	4
Cas9 protein	+	+	+	-
gB2m.1.3	+	+	-	-
ssODN.B2m.SilentM	+	-	+	-

B



C

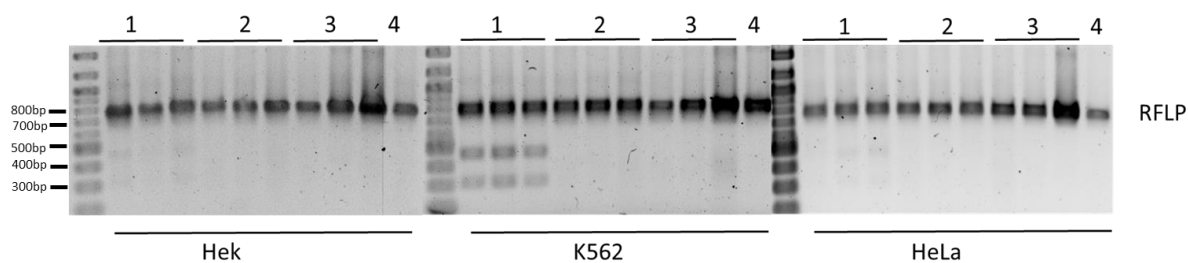


Figure 5.11. Evaluation of HDR in HeLa, HEK and K562 cell lines.

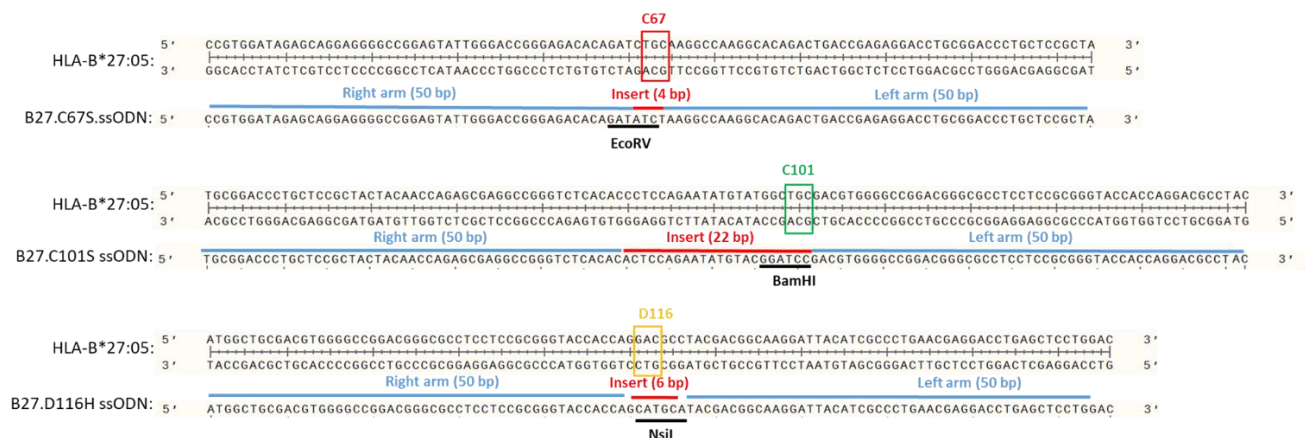
DsODN donor templates, ssODN.B2M.SilentM and RNP.B2M-1.3 were transfected into HeLa, K562 and HEK cells. (A) Table summarizes the components of each transfection. (B,C) At one day post-transfection, PCR amplicons of transfected cell-derived genomic DNA were used to check the cleavage and HDR efficiency by T7EI and RFLP assays, respectively. The size of the wt B2M amplicon was 762bp and the sizes of T7-cleaved products were 458 and 304 bp. The sizes of RFLP products were 455 bp and 307 bp. Data are presented of one experiment.

5.2.6 HDR of C67S, C101S and D116H using gRNA and ssODN

After optimized sgRNAs had been determined for introducing DSBs on or near the target regions, I designed donor templates to enable homologous recombination and insertion of the desired mutations into the HLA-B27 sequence. Besides nucleotide substitutions at C67, C101 and D116, additional silent mutations, which was used in SDM experiments (Figure 5.1), were also included in the designed donor templates. The reason for employing silent mutations was to create a new restriction enzyme site to evaluate HDR efficiency and screening of single-cell clones. All donor templates were ordered in a form of single-stranded donor oligonucleotides (ssODNs) which were suitable for short insertions and could provide high HDR efficiency compared to similar dsDNA donor templates [367, 368]. The criteria for designing donor templates included: (1) containing targeted codon changes to be introduced along with silent mutations; (2) have 40-60bp homology arms; (3) Avoid local GC content extremes and secondary structures of the target sequence.

By applying the above criteria, I designed three separate ssODN donor templates for introducing the following genetic modifications, C67S, C101S and D116H, into the HLA-B*27:05 allele (Figure 5.12). Beside amino acid substitutions, ssODN of S67, S101 and H116 also contained the additional silent mutations for introducing the restriction enzyme sites of EcoRV, BamHI and NsiI, respectively. While Both B27.C67S.ssODN and B27.D116H.ssODN have nucleotide substitutions close to the 5' end of crRNA and PAM motif of crRNA-target sequences, nucleotide substitutions of B27.C101S.ssODN were located at the 3' end of crRNA. To decrease the possibility of Cas9 RNP-mediated cleavage within the donor template, an extra amino acid substitution was added to change the PAM motif (CCC to CAC) and prevent sgRNA

binding. Both right and left homology arms of all ssODNs were 50 nt in length. Donor templates were ordered from IDT in the form of single-stranded DNA oligos (Alt-R HDR Donor Blocks).



Amino acid change	sgRNA name	ssODN name	ssODN length (nt)	RE site	Cleaved bands (B27 primer 4)
67.C>S	sg272F	B27.C67S.ssODN	104	EcoRV	441 + 472 (bp)
101.C>S	sg359R	B27.C101S.ssODN	122	BamHI	339 + 574 (bp)
116.D>H	sg426F	B27.D116H.ssODN	106	NsiI	294+619 (bp)

Figure 5.12. Design of donor templates built for homology-directed repair to introduce single amino acid changes within the HLA-B27 sequence.

All donor templates were designed and produced in the form of ssODN. Each ssODN contained an insert sequence (red), homology arms (blue) and restriction enzyme sites (black). The table summarises the details of each ssODN.

Firstly, I aimed to test the CRISPR/Cas9 system and donor template in cell lines expression B27 cDNA, then move to targeting genomic DNA sequences. For HDR of HLA-B27, ssODN donor templates, RNPs of Cas9 protein and sgRNA were co-

transfected using electroporation with the Amaxa Nucleofector system (Lonza). Equal numbers (10^6) of HeLa-B27 cells were transfected with ssODN.D116H and RNP complexes (of Cas9 nuclease and sg426F) at a molar ratio of 4:1 followed by incubation in complete culture media with either HDR Enhancer, DMSO or media alone. Genomic DNAs were extracted from post-transfected cells and used as templates for producing PCR amplicons for T7EI and RFLP assays, which were applied to evaluate the cleavage and HDR efficiency, respectively. The T7EI assay (Figure 5.13) indicated that Cas9/sg426F RNP complexes can introduce DSBs in all samples as indicated by the generation of the bands 619 bp and 294 bp. However, the RFLP assay did not demonstrate cleavage of the amplicon. Therefore, even in the presence of B27.D116H.ssODN donor template, DSBs were mainly repaired following the NHEJ pathway by HeLa-HLA-B27. Similar results were observed with repeated experiments using sg and ssODN C67S (sg272F and B27.C67S.ssODN) and C101S (sg359R and B27.C101S.ssODN) pairs (data not shown).

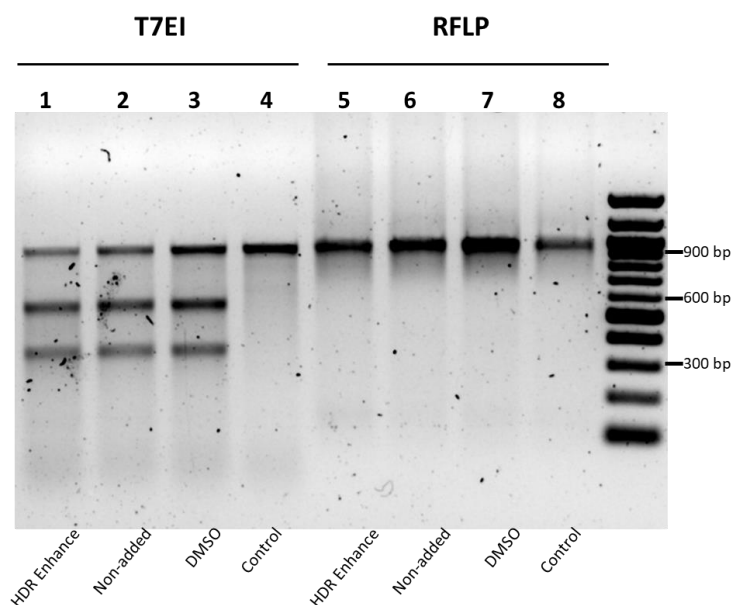


Figure 5.13. Genome editing by transfection of B27.D116H.ssODN and Cas9/sgRNA RNP targeting the HLA-B27 cDNA expressed by HeLa-B27 cell lines.

Cells were transfected with the complete CRISPR/Cas 9 platform (ssODN + Cas9/sgRNA RNP complexes) or Cas9 only as a negative control (Control). After transfection, cells were grown in cDMEM or cDMEM containing HDR Enhancer or DMSO. PCR amplicons of transfected cell-derived genomic DNA were used to check HDR efficiency by T7EI (lane 1-4) and RFLP assay (lane-4). Experiments have been repeated three times using three different pairs of sgRNA and ssODNs. Data are presented of one experiment.

In section 5.24, K562 cells had been shown to be more suitable for gene-insertion than HeLa cell lines. Therefore, I repeated the HDR experiments with K562-HLA-B27 cell line which was kindly provided by Dr. Simon Powis (St. Andrew University). Equal number (10^6) of K562-HLA-B27 were transfected with Cas9 RNPs and corresponding ssODNs using electroporation. Cells transfected with the Cas9 protein only were used as negative controls. DNA gel electrophoresis of the T7EI assay (Figure 5.14A) demonstrated cleavage products as detected in lanes 1, 2 and 3. This result indicates that RNP complexes of Cas9 nuclease and sgRNAs can introduce DSBs within the HLA-B*27:05 gene expressed by K562-HLA-B27 cells. Following the RFLP assay, PCR amplicons were incubated with the corresponding RE for each target including EcoRV for C67S, BamHI for C101S and NsiI for D116H. The RFLP assay (Figure 5.14B) demonstrated BamHI (lane 5) and NsiI (lane 6) cleavage products within the negative controls. These cleavage bands were digest products resulting from the multiple cloning region of the pCR3 plasmid which was the backbone plasmid used to integrate HLA-B27 cDNA into the genome of K562 cells. Cleaved bands were observed in samples transfected with RNP complexes and ssODNs (lane 1, 2 and 3). The RFLP assay of gene-edited cells indicated that K562-HLA-B27 cells have repaired DSB following HDR, which facilitated the introduction of the desired mutations.

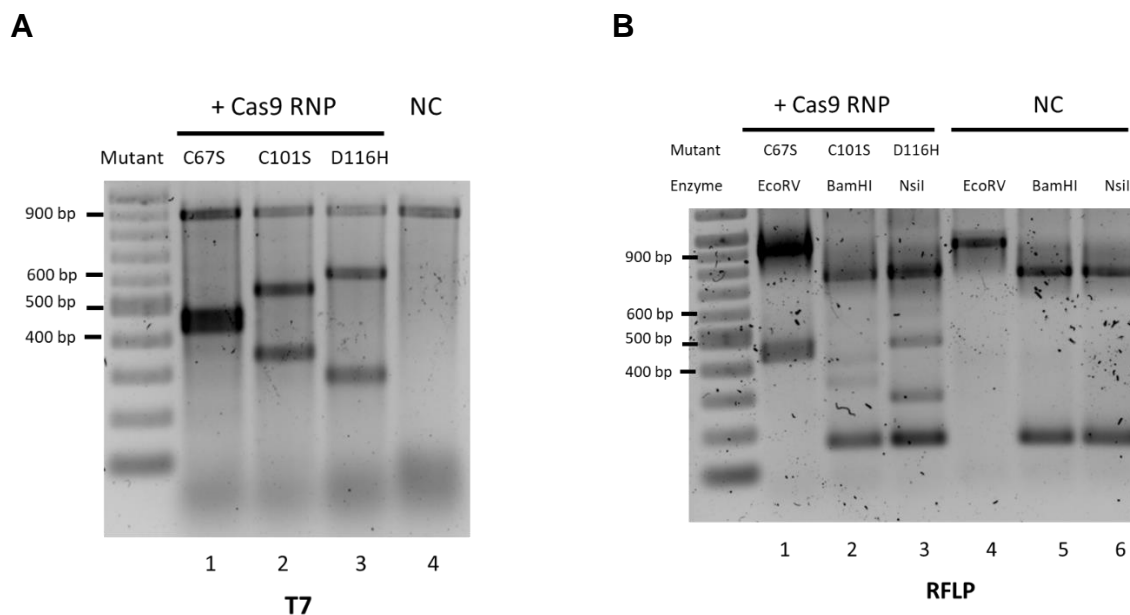


Figure 5.14. Genome editing by transfection of ssODNs and Cas9/sgRNA RNPs targeting HLA-B27 in K562-HLA-B27 cell line.

Cells were transfected with a completed platform composed of ssODN + Cas9/sgRNA RNP complexes or Cas9 only as negative control (NC). After transfection, cells were grown in cRPMI containing HDR Enhancer. PCR amplicons of transfected cell-derived genomic DNA were used to check HDR efficiency by T7E1 (A) and RFLP assays (B). Data are presented of one experiment.

5.2.7 Cloning of gene-edited cells

After achieving HDR at the target sites of the HLA-B27 cDNA, clonal cell lines were generated from polyclonal pools of gene edited cells by single-cell cloning. Single-cell clones of gene-edited K562-HLA-B27 were expanded and partly harvested for extraction of genomic DNA. PCR amplicons of each single-cell clone using F4-B27 and R4-B27 were used to identify single cell clones containing desired mutations by the RFLP assay. Electrophoresis results of the RFLP assay (Figure 5.15) identified 3 in 6 screened clones carried the C67S mutation (50%), 2 in 10 screened clones carried

the C101S mutation C101S (20%), and 2 in 13 screened clones carried the D116H mutation (15.4%).

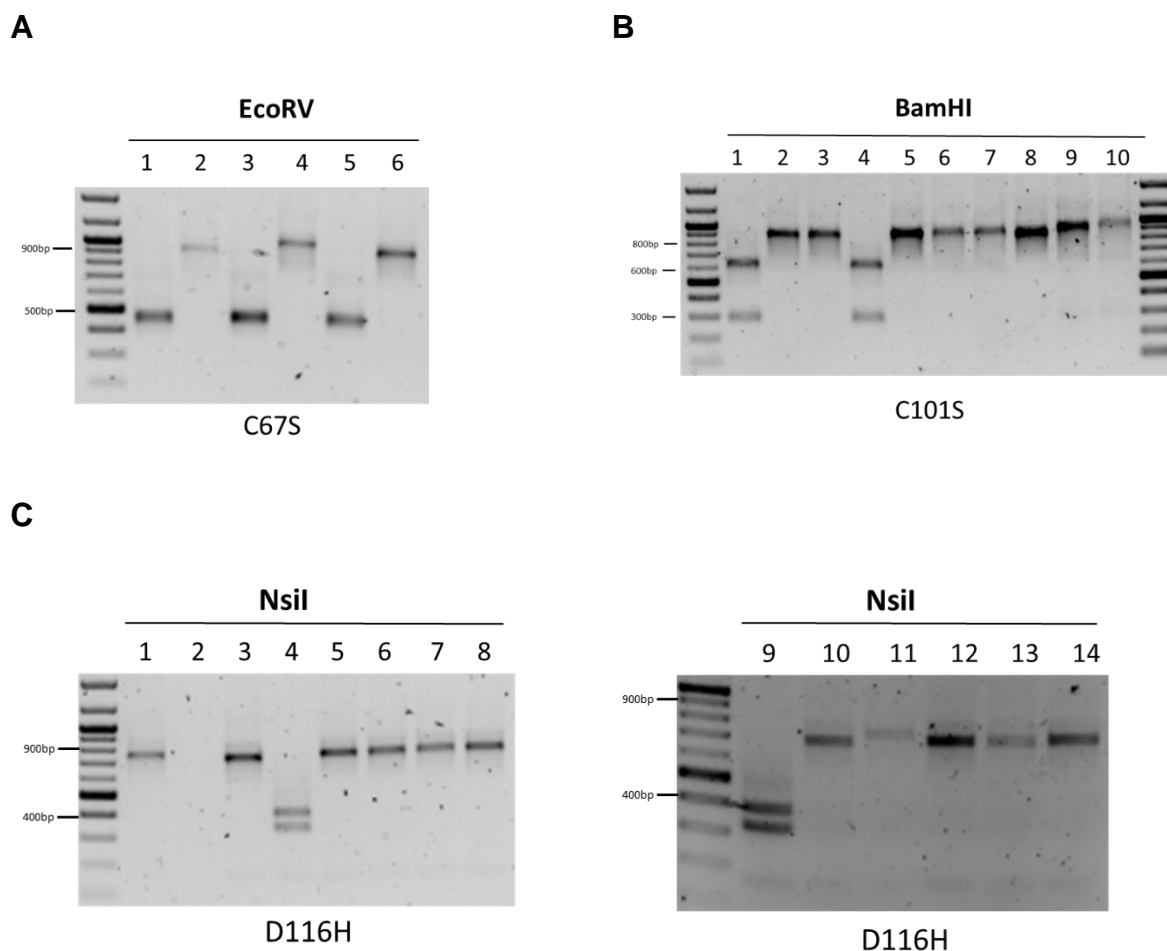


Figure 5.15. Single-cell isolation of monoclonal K562-HLA-B27 cells containing target mutations.

Monoclonal cells were isolated using serial dilution of polyclonal populations. Single-clonal cells were identified using the RFLP assay with corresponding RE. (A) PCR amplicons of single-cloned K562-B27-C67S cells were digested by EcoRV. Undigested products were observed in lane 2, 4 and 6. Cleaved products (441 + 472bp) were observed in lane 1, 3 and 5. (B) PCR amplicons of single-cloned K562-B27-C101S cells were digested with BamHI. Undigested products were observed in lane 2, 3 and 5-10. Cleaved products (339 + 574bp) were observed in lane 1 and 4. (C) PCR amplicons of single-cloned K562-B27-D116H cells were digested by NsiI. Undigested products were observed in lane 1, 3, 5-8 and 10-14. Cleaved products (294 + 619bp) were observed in lane 4 and 9.

RFLP-mediated screening of single cell clones revealed clones containing the desired mutation and were expanded for confirmation by Sanger sequencing before storage and use for downstream experiments. PCR amplicons were prepared using the F4-B27 and R4-B27 primer pair and sequenced using primer F4-B27 (Appendix 5). The sequencing chromatograms of wild-type and mutated HLA-B27 are shown in Fig 4.19. Sequencing confirmed the mutations were introduced and that the genotypes were homozygous. Therefore, CRISPR-mediated HDR experiments were successful in introducing different mutations into the HLA-B27 cDNA sequence expressed in the K562 cell line.

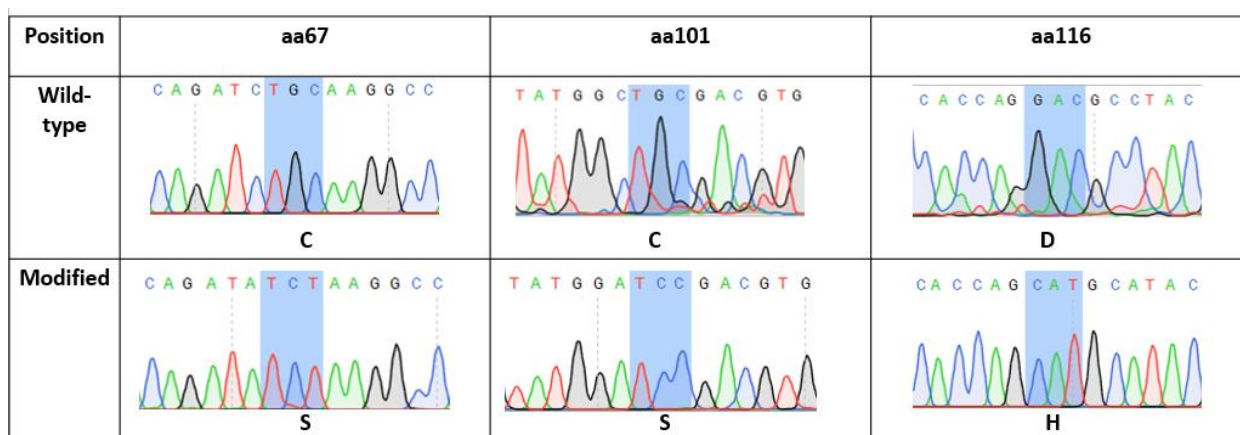


Figure 5.16. Confirmation of nucleotide substitutions in monoclonal gene-edited K562-HLA-B27 cells.

HLA-B27 short segments of wild-type and gene-edited K562-HLA-B27 were shown in the top and bottom rows, respectively. The nucleotide missense mutations were highlighted in blue bars. The HLA-B27 sequence of K562-B27-C67S contains nucleotide substitutions from TCT to TGC. The HLA-B27 sequence of K562-B27-C101S contains nucleotide substitutions from TCC to TGC. The HLA-B27 sequence of K562-B27-C101S contains nucleotide substitutions from CAT to GAC.

5.2.8 Investigating gene edited HLA-B27 expression

To investigate how CRISPR-mediated nucleotide substitutions affected the cell surface expression of HLA-B27, each clonal gene-edited K562 cell line was analyzed by flow cytometry using W6/32 and ME1 antibody. Flow-cytometry data of each sample stained by each antibody were compared to K562.B27 samples using ordinary one-way ANOVA (Prism) (Appendix 4, Supplementary Data 14 and 15). Flow cytometric analysis (Figure 5.17) revealed that ME1- reactive HLA allotypes were not expressed by K562 cells (MFI: 192 ± 2). The surface expression of HLA-B27 on K562-B27-C101S (MFI: 261 ± 36) were significantly lower than that of K562 (MFI: 7262 ± 295). W6/32-reactive HLA allotypes were also observed to be decreased in K562-B27.C101S compared to wild type K562. These results suggest that CRISPR/Cas9 RNP targeting position 101 may not only affect the HLA-B27 allele but also other HLA class I sequences. To verify these potential off targeting effects, deep-sequencing should be performed in the future to assess indel mutations within non-B27 HLA class I alleles to identify the presence of off-target mutations.

The fully assembled ME-1 reactive HLA-B*27:05.C67S heavy chain was expressed at lower levels than wild type at the cell surface (HLA-B27.C67S 4555 ± 117 compared to WT HLA-B27 7262 ± 295). However, expression of B27.C67S was significantly higher than B27.C101S. There appears to be no significant difference in cell surface expression levels between K562-HLA-B27 (W6/32: 5811 ± 72 , ME1: 7262 ± 295) and K562-B27.D116H (W6/32: 5816 ± 70 , ME1: 6798 ± 258). Hence, the amino substitution at p116 within the F pocket of the peptide binding groove has no impact on surface expression of HLA-B27 heavy chain.

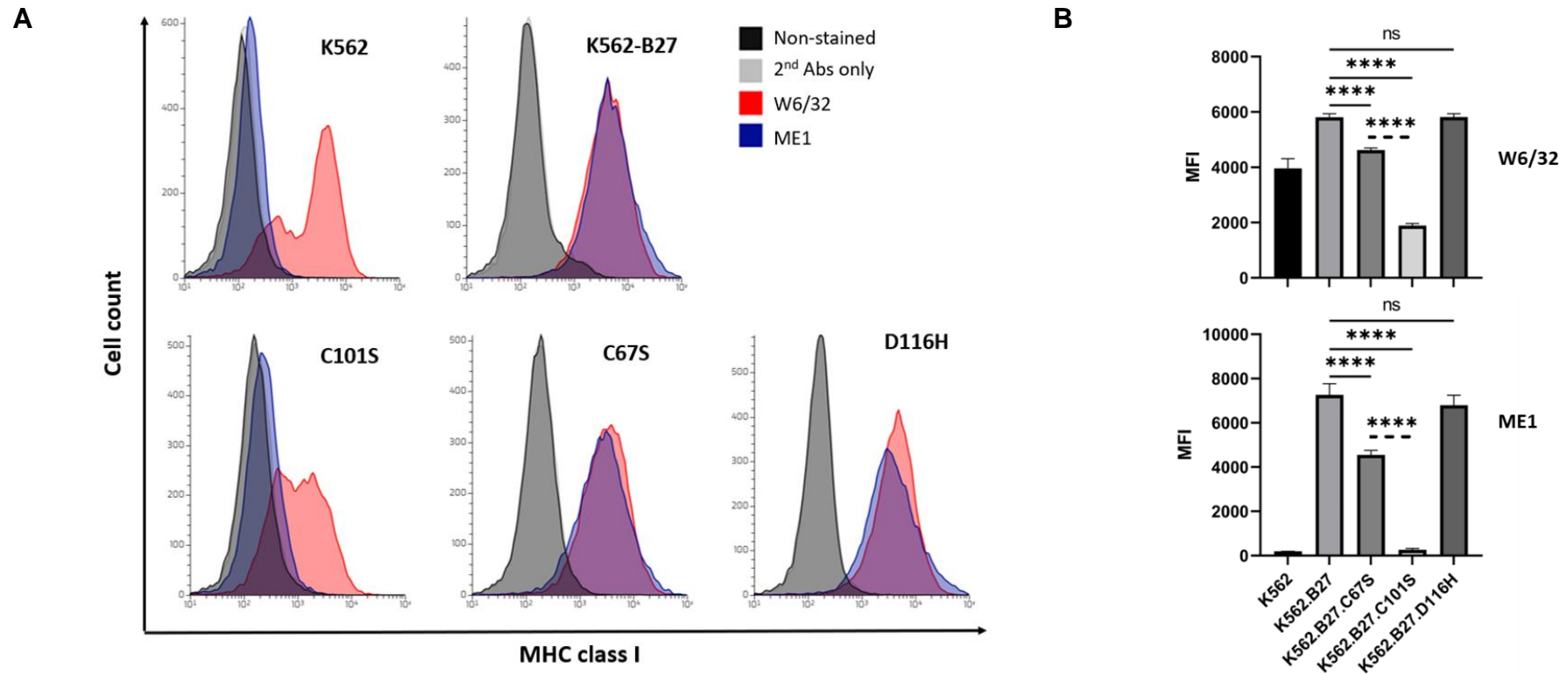


Figure 5.17. Characterisation of surface MHC-I expression by gene-edited K562-HLA-B27 cells.

Cell surface expression of MHC-I and HLA-B27 by K562 cells and gene-edited K562 cells stably expressing HLA-B*27:05 (wild type), C101S, C67S and D116H as determined by flow cytometry using W6/32 and ME1 antibodies. (A) Flow cytometry data was presented as histogram plots of cell counts versus Mean Fluorescence Intensities (MFI) corresponding to the expression of MHC-I or HLA-B27. (B) The bar charts demonstrate the mean MFI data (mean \pm SEM, n=3). One-way ANOVA with Turkey's post-test. ****P<0.001.

I then wanted to determine the folding status of HLA-B27 and the respective mutants with respect to HC-dimerization. Equal number (10^6) of cells were pre-treated with NEM followed by lysis in 1% NP-40 detergent. Non-reduced and reduced lysates were resolved on 10% SDS-PAGE followed by immunoblotting with HC-10 antibody and GAPDH as a protein loading control. Immunoblotting with GAPDH (Figure 5.18, bottom blot) indicates that there was approximate equal protein loading between cell lysates. Immunoblotting with HC10 of reduced cell lysates (Figure 5.18, middle blot) detected a single MHC-I isoform (~40 kDa). Immunoblotting with HC10 of non-reduced cell lysates (Figure 5.18, top blot) demonstrated only one monomer (vi) and different isoforms of HLA-B27 heavy chains ranging between 75 to 110 kDa. Immunoblotting detected 3 isoforms of homodimers (arrow ii-iv) for both HLA-B*27:05 (lane 2) and B*27:05-D116H (lane 6) heavy chains. These observations suggest that substitution to the HLA-B*27:09 subtype did not affect the HC-dimerization of HLA-B27 by K562 cells. Interestingly, substitution of cysteines at p67 and p101 can lead to different levels of HLA-B27 expression as well as differences in detection of homodimer isoforms. K562-B27-C67S (lane 3) demonstrated apparent lower HLA class I heavy chain protein levels as detected by HC10. On the other hand, HLA-B27-C101S (lane 4) demonstrated enhanced protein levels following immunoblotting of reduced lysates and the detection of 5 different isoforms of HLA-B27 HC-dimers (i-v). Therefore, a combination of flow cytometry analysis and immunoblotting suggests that both cysteine mutants at p67 and p101 reduce HLA-B27 expression at the cell surface but for different reasons. While C101S led to insufficient folding of HLA-B27 molecules and gave rise to different types of aberrant heavy chain molecules, C67S led to a reduction in the expression level of intracellular HLA-B27 molecules.

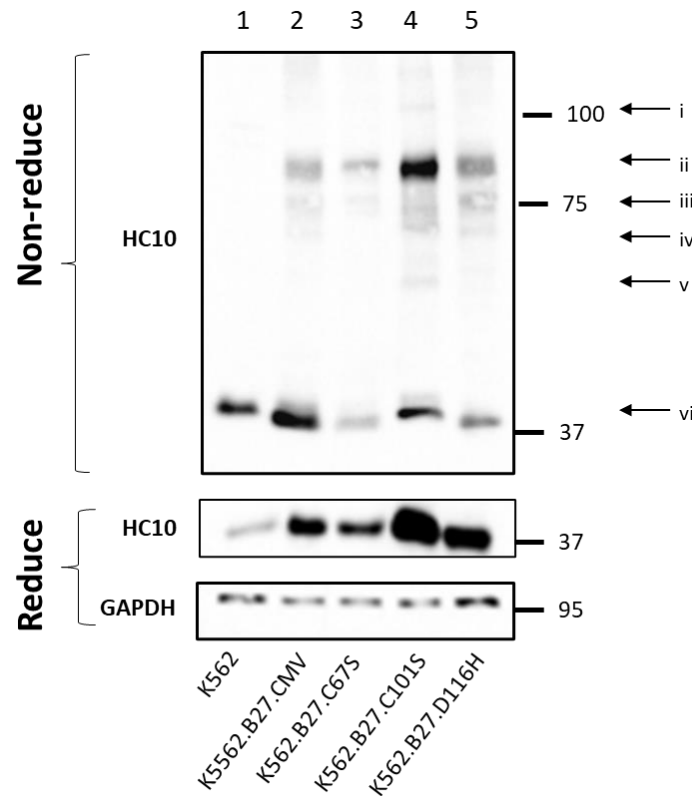


Figure 5.18. Immunoblotting of HLA-I heavy chain expressed by gene-edited cells.

K562 cells and gene-edited K562 cells stably expressing HLA-B*27:05 (wild type), C101S, C67S and D116H were lysed in 1% NP-40 detergent. Non-reduced and reduced lysates were resolved on 10% SDS-PAGE followed by immunoblotting with HC10 antibody (anti- unfolded MHC-I). Reduced lysates were resolved on 10% SDS-PAGE followed by immunoblotting for GAPDH as an internal protein level control.

5.2.9 Evaluating off-target effects on gene-edited cell lines

Undesirable off-target mutations were screened by NGS at genome regions which have high levels of homology to the targeted sequence. Deep sequencing was applied to evaluate the off-target effects caused by the Cas9 RNP of sg272F, sg359R and sg426F. Firstly, potential off-target sequences were predicted using COSMID, an online software tool for identifying and validating CRISPR off-target regions [353]. The criteria for identifying off-target sites included: (i) screening whole human genomes; (ii) containing maximum of 3 mismatches and 1 indel mutation; (iii) <2.5 in score of off-target sites for the input target sites. Lower ranking score of off-target sites indicates the more likely off-target sites. Following the above criteria, only one locus was identified as a potential off-target site by Cas9 RNP of sg272F. The potential off-target loci, named as OT-sgB27.272_2, are located within the *FLJ11042* or *ACOXL* gene on chromosome 2. The on-target region of sg272F was also included in deep sequencing experiments to investigate the percentage of indel mutations.

Table 5.1. On-target and off-target sites for the RNP of Cas9 protein and sg272F.

Off-target regions of sg272F were predicted by COSMID software. Mismatch or indel mutations are highlighted in yellow. Query type indicates the type and number of changes examined. Off-target ranking scores are based on the number and location of base mismatches.

Name	Result	Query type	Mismatch	Chr Position	Strand	Score
OT.sgB27.272_1	CCGGGAGACACAGATCTGCAGGG -- hit CCGGGAGACACAGATCTGCANGG -- query	No indel	0	Chr6		0
OT.sgB27.272_2	CCTGGAGACACAGCTCTGCAGGG -- hit CCGGGAGACACAGATCTGCANGG -- query	No indel	2	Chr2:111031770-111031792	+	1.45

Primers for NGS were designed to produce short PCR amplicons with the size of 200 to 250 bp (Appendix 5). Genomic DNA of wt K562-HLA-B27 cells and polyclonal gene-

edited K562-HLA-B27 cells, which were transfected with the RNP of Cas9 protein and sg272F, were extracted and used as templates for preparing the initial amplicons. After generating libraries with unique bar codes for each sample, PCR products were subsequently sequenced on the Illumina MiSeq at 100,000 reads per sample.

Off-target effects were evaluated based on the comparison between NGS sequencing data and reference genome sequences. The percentage of variants or indel mutations was analyzed through re-alignment programs Geneious by the following criteria: (i) identifying variants within the entire range covered by the reads; (ii) eliminate variants with low frequency (<0.5%). Off-target analysis (Figure 5.19) indicated that the on-target site contained 63.1% (n=2) indel mutations generated during DSB repair following NHEJ pathway. More importantly, the percentage of indels detected in potential off-target regions of gene-edited cells were very low at 0.3% (n=2). These percentages and types of variants were the same as that of untreated K562-HLA-B27 cells. Therefore, off-target analysis indicated that indels were also not detected in the screened off-target locus of sg272F.

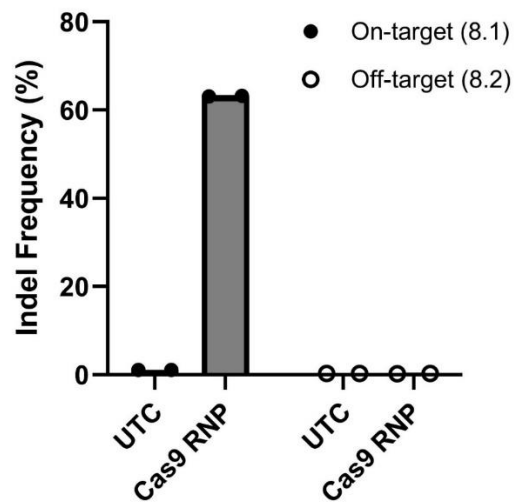


Figure 5.19. On-target and off-target analysis of the RNP of Cas9 protein and sg272F on K562-HLA-B27 cells

Indel frequencies for on-target and potential off-targets identified by Miseq deep-sequencing in untreated K562-HLA-B27 cells (n=2) and CRISPR/Cas9-treated K562-HLA-B27 cells (n=2).

5.3 Discussion

To the best of my knowledge this is the first study exploiting the CRISPR/Cas9 system to directly introduce nucleotide substitutions within HLA class I genes. The targeted gene was the HLA-B*27:05 subtype which is strongly associated with Ankylosing Spondylitis (AS) [174]. CRISPR/Cas9-mediated modification of HLA-B*27:05 has the potential to be developed into a therapeutic treatment for AS in the future.

The aim of this chapter was to test the idea of manipulating protein folding by introducing different nucleotide changes into the HLA-B*27:05 sequence and investigate how misfolding events and potentially how the UPR were influenced by these variants. Three amino acid changes were targeted at p67 (C67S), p101 (C101S) and p116 (D116H).

Cys⁶⁷ has been implicated in the dimerization process especially cell surface dimers [369] and B27.C67S transgenic rats develop disease with a milder phenotype than wild-type B27 transgenic rats [180]. On the other hand, D116H will transform HLA-B*27:05 to HLA-B*27:09 which could improve the protein folding process. A C101S mutant was generated as a control to assess the effects of a fully misfolded HLA-B27 molecule.

The tendency for the HLA-B27 heavy chain to misfold and/or form homodimers has been postulated to contribute to AS pathogenesis [370]. Using two different methods which are site-directed mutagenesis and CRISPR/Cas9, the cell models containing C67S or C101S or D116H were generated regarding the dimerization and misfolding events of B27 molecules of each different population. The monomer of B27 molecules have three different isoforms which can be explained by 3 different states including

non-reduced (high mw), partly-reduced (medium mw) and oxidized (low mw). The three high mw protein bands (90-130kDa) maybe representative of different homodimer conformers of B27 molecules or heterodimers of HLA-B27 and other proteins related to folding or the antigen-presenting pathway.

In both the SDM and gene-edited cell models, the absence of Cys¹⁰¹ facilitated high levels of HLA-B27 dimerization and prevented the HLA-B27 molecule being expressed at the cell surface. The absence of Cys⁶⁷ can still form high molecule weight disulphide bonded structures but at lower expression level and exhibit an altered number of B27 heavy chain isoforms. This result proposed that modification of C67S can potentially reduce the misfolding events and still maintain the presence of HLA-B27 heavy chain on cell surface.

HLA-B*27:05 and -B*27:09 are AS and non-AS associated subtypes, respectively which differ at a single residue at position 116 within the heavy chain. The presence of Asp (D) in HLA-B*27:05 has been found to play an essential role in AS susceptibility [371]. I transformed HLA-B*27:05 to 09-like HLA-B*27:05 by replacing Asp¹¹⁶ by His using CRISPR/Cas9 system. With both SDM and gene-edited cell models, there were no differences in either expression levels and HLA-B27 heavy chain misfolding as determined by HC-dimerisation. These observations suggest Asp¹¹⁶ may not be involved in HC-dimerisation. Previous studies by Cheng [371] and Rana [362], comparing HLA-B*27:09 and 05, reported the latter to strongly bind to more self-peptides which could potentially become auto-antigen targets. To determine whether there are differences in the peptide pools, analysis of the presented peptides on HLA-B*27:05 and 09 should be performed. Additionally, it is also necessary to examine the maturation rates of HLA-B27 heavy chains within the gene edited cells.

In this chapter, I developed an efficient CRISPR/Cas9-mediated system that precisely altered nucleotides within the HLA-B27 cDNA sequence. The CRISPR/Cas9 system consists of a short ssODN donor template and a RNP composed of a Cas9 nuclease and sgRNAs targeting the HLA-B27 sequence. A major challenge for precise CRISPR/Cas9-mediated knock-in is the ability to diminish NHEJ background and induce high-fidelity HDR edits. In this study, HDR efficiency was observed to be very low in HeLa compared to HEK and K562 cells. Although different Cas9 systems, including Cas9 nickases, catalytically dead Cas9 fused to FokI and transcription activator–like effector nucleases, have been applied in HeLa cells, none can induce HDR within this cell line [372]. The variable ratio of HDR and NHEJ between cell types can be explained by different epigenetic modifications [373] and components of the DNA repair machinery.

The next stage of study will be to apply this CRISPR/Cas9 system to target the genomic HLA-B*27:05 allele. EBV immortalized B cells containing wild-type HLA-B27 allele are used to optimize the CRISPR/Cas9 system before applying on iPSC or HSCs. The current systems were designed to target both cDNA and genomic DNA of the HLA-B*27:05 allele, but the donor template should be modified to encompass the intronic regions within the homology arms.

HLA class I heavy chain expression has previously been successfully disrupted to enhance the immunological compatibility of iPSC cells [315]. With the effective CRISPR-mediated switch in HLA subtype reported in this study and combined with reported strategies enhancing immunological compatibility, this approach could greatly increase donor compatibility. Another potential application of HLA modification is in the treatment of HLA-associated diseases such as AS. In 2012, the first study

using autologous hematopoietic stem cells (HSCs) for treating ankylosing spondylitis were reported by Britanova *et al* [374]. Recently, CRISPR-mediated gene editing HSCs has been approved by the FDA for treatment of genetic blood diseases, including sickle cell disease [375] or β -thalassemia [376]. These clinical trials have started new promising approaches which could be applied in the treatment of severe autoimmune diseases using HSC transplantation.

In summary, this is the first study reported using the CRISPR/Cas9 system for the modification of HLA-B27 heavy chain to introduce single amino changes. My system, consisted of Cas9 RNP and donor template, precisely modifying the human HLA-B*27:05 sequence and successfully facilitating insertion of desired mutations into the target region. Off-target analysis indicated that indels were also not detected at the screened off-target locus of sg272F.

Data indicates that the C67S mutation of HLA-B*27:05 has led to a slight reduction in the expression of this heavy chain at the cell surface and reduced misfolding events of HLA-B27 molecules. On the other hand, HLA-B*27:05 and 09 exhibit similar levels of homodimers as well as cell surface expression. For future work, peptidome analysis and maturation analysis can be considered to determine the full extent of the differences between HLA-B*27:09 and 05 and perhaps reveal potential mechanisms of association/non-association with AS. When the AS disease mechanism is elucidated, my study may form the foundation for applying CRISPR/Cas9 technology to treat AS.

CHAPTER 6
CONCLUSIONS

Chapter 6 Conclusions

6.1 General discussion

Thus far, this study demonstrated successful applications of CRISPR/Cas9-mediated knock-in with respect to modifying MHC-I – peptide complexes. The first of these involved modifying the actual peptide cargo which was achieved by initially evaluating the expression of PLB constructs in B2M-deficient cells (1.1). By applying the data from the PLB analysis and developing the methodology, exogenous peptides could be potentially loaded into MHC-I by CRISPR-Cas9; (1.2). Furthermore, I employed CRISPR-Cas9 to introduce mutations within HLA-B*27:05 and evaluate their biochemical characteristics. These above studies were assessed for their potential off-target effects by deep sequencing to evaluate their safety as possible therapeutic tools.

Within chapter 3 I had established the first PLB containing peptides derived from SARS-CoV-2 virus. Both PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ and PLB-S₂₆₉₋₂₇₇-G₆(SGG)₃ proteins were found to rescue the expression of surface MHC-I, especially their antigen-specific HLA i.e. HLA-A*02, in B2M-deficient cells. These PLB constructs were the premise for the development of a CRISPR-mediated system that precisely inserted the peptide and linker into the human *B2M* locus. In chapter 4, optimized CRISPR systems were developed which included the RNP of Cas9 nuclease and sgRNAs targeting exon 1 of the *B2m* locus, along with megamer donor templates containing sequences of peptide, linker and homology arms. While most CRISPR studies normally use short donor templates to facilitate the insertion of a few nucleotide changes, I exploited the megamer donor template to knock-in long DNA sequence (about 75 nucleotide) into the target-site. Although the N₃₂₅₋₃₃₃-(GGGGS)₃ sequence was successfully integrated

into the endogenous *B2M* allele, the PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ molecules were not expressed at the expected size by gene-edited cells. Based on the cDNA analysis, the lack of PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ was explained by the intron-exon splicing process as the linker-peptide sequence had introduced intronic motifs which could lead to the removal of peptide and linker sequences from the mRNA of the overall PLB sequence. Thus, intronic sequences should be taken into consideration and location of insertion when designing donor DNA templates. Additionally, it is necessary to develop techniques to identify the precise intron-exon splice sequences among several possible sites and remove unwanted sequences.

Using the information and restrictions from attempts to knock-in the PLB-N₃₂₅₋₃₃₃-(GGGGS)₃ sequence, all the intron motifs within the S₂₆₉₋₂₇₇-G₆(SGG)₃ donor template were replaced by silent mutations before successful knock-in of the target region was achieved. The PLB-S₂₆₉₋₂₇₇-G₆(SGG)₃ proteins was found to be expressed with the expected size and had assembled with endogenous HLAs which appear to be presented at the cell membrane. In conclusion, specific-peptides can be introduced into B2M via CRISPR/Cas9 to produce PLB molecules which facilitate MHC-I peptide presentation.

With the special characteristics of PLB, this technology can be applied in the biomedical field. In previous studies, the SCT platform has been proposed as a composite vaccine targeting multiple aspects of immunity against viruses such as *West Nile* or *Listeria monocytogenes* virus [189, 377]. Compared to SCTs, the PLB protein is smaller in size due to the lack of heavy chain, therefore, it is more convenient for packaging into vector plasmids as well as reducing the toxicity for cells during delivery. Based on precise algorithms, NetMHCpan or IEDB can predict pathogenic

peptides for each HLA type. Taking advantage of these algorithms, the design of PLB platforms carrying desired peptides, of proposed high affinity and specific to endogenous patient HLA-I, increases the feasibility of developing personalized vaccines. In the case of low-affinity peptides, CRISPR-mediated knock-in of PLB can possibly generate stable expression of PLB on APCs such as DCs. The use of DCs expressing PLB is a promising approach to presenting antigens to activate tumor-specific T cells.

The chapter 5 in this thesis I demonstrated the use of CRISPR/Cas9-mediated modification of HLA-B*27:05. All three amino acid changes; C67S, C101S and D116H were successfully introduced into the cDNA of HLA-B*27:05 using the CRISPR/Cas9 system with different sgRNA and corresponding ssODN donor templates. The C67S and D116H mutations were introduced with a view for future potential therapeutic development. The C101S mutation was used as a control for introducing a misfolded version of HLA-B27. Whilst the introduction of C67S within HLA-B27 can potentially reduce misfolding events and still maintain the presence of this heavy chain on cell surface, that of D116H switches HLA-B*27:05 to the HLA-B*27:09 subtype, which is a non-AS-associated HLA allele. This is the first study using the CRISPR/Cas9 system for the modification of HLA class I heavy chains and opens a promising approach for applying CRISPR/Cas9 technology in manipulating immune responses.

Before translating the manipulation from bench to bed, it is necessary to evaluate the feasibility of therapy on HLA-B27 transgenic rat model. Previous studies have demonstrated that healthy rat receiving engraftment of bone marrow cells from HLA-B27 transgenic rat can develop AS-like conditions [116]. Thank to recent advancement in gene-editing, the genetics of bone marrow cells or HSCs can be modified before

transplantation [378]. Based on that, the CRISPR-mediated C67S and D116H modification can be applied on HSCs of HLA-B27 rat and subsequently transplant the gene-edited engraftment back to evaluate the manifestation of AS. If the disease is driven by misfolding hypothesis, the recovery or slow disease progression is expected to be observed on HLA-B27.C67S rat model. On the other hand, D116H modifications which introduce the changing in the peptide binding pocket of HLA-B27 molecule, is expected to prevent the presentation of AS-related arthritogenic peptide and reduce autoimmune responses. Evaluation of immune responses is also extremely critical because genetic modification of HLA class I also bring the risk of generating novel T cells responses and potentially triggering NK cells.

6.2 Limitations and future work

Although either single base substitutions or long nucleotide sequences have been successfully inserted into genomic DNA via CRISPR/Cas9 technology in this study, there are still changes that can be applied to reduce the off-target effects and enhance HDR efficiency by altering delivery options or Cas9 variants. Firstly, a combination of Cas9 RNP and AAV serotype 6 for donor template delivery has led to efficient HDR in human T cells and hematopoietic stem and progenitor cells (HSPCs) [293, 379]. Therefore, the CRISPR-Cas9/AAV6 approach can be applied to introduce PLB into the *B2M* locus or nucleotide substitutions into HLA alleles in patient-derived HSPCs. Second, to overcome the concern of off-target effects, different versions of sgRNAs and Cas9 nuclease has been developed. For instance, sgRNA with two additional G nucleotides at the 5' end has been demonstrated to discriminate off-target sites effectively [352]. Another strategy to minimize off-target activities is to use the high-fidelity SpCas9 mutant with a single point mutation, p.R691A [380]. As mentioned in 1.2.2.1, Cas9 nickase, containing mutations within the nuclease domains RuvC (D10A) or HNH (H840A), only induce single-stranded breaks which are more favorable for repair by the base excision repair pathway so it can lead to fewer indels than the original Cas9 nuclease [381]. Most recently, Chu and his colleagues have established a “spacer-nick” CRISPR/Cas9^{D10A} system to improve suitability and safety for gene therapy [382]. This system was developed based on a combination of Cas9^{D10A} nickase, a pair of PAM-out sgRNA at long (>200 bp) spacer distances and an AAV6 donor template. The spacer-nick system has been applied to repair disease-related mutations occurring within the *ELANE*, *HBB*, *IL7R* and *PRF1* genes, the results of

which have demonstrated HDR efficiencies of 20-50% with minimal indel mutations at the target site.

To minimize the components needed to be delivered for CRISPR/Cas9-mediated gene-editing purposes, many groups have fused different proteins into the Cas9 nuclease to achieve precise DNA changes without requiring DSBs or exogenous donor DNA repair templates. For example, a single-stranded DNA deaminase enzyme has been fused with an impaired Cas nuclease (which cannot generate DSBs) to form new platforms called base editors including adenine base editor (ABE) and cytosine base editor (CBE). ABEs and CBEs can install four transition mutations ($T \rightarrow C$, $G \rightarrow A$, $C \rightarrow T$, $A \rightarrow G$) and have been applied to introduce DNA modifications in different cell types and organisms [383]. However, the platform cannot perform eight transversion mutations such as $A \rightarrow T$ or $C \rightarrow G$ which would have been required in this study when introducing the respective C67S and D116H mutations within HLA-B*27:05. Liu's study [314], used sgRNAs which were engineered to contain a sequence used as a template for reverse transcription (RT) and was known as a primer editor sgRNA (pegRNA). The RT template of pegRNA was complementary to 5' non-editing DNA and works as a primer for DNA elongation. The newly synthesized single-strand DNA (ssDNA) contained edited sequences which were ligated to the parental DNA to form a heteroduplex. DNA repair mechanisms would convert the heteroduplex to the permanently edited DNA. The PE approach exhibited successful gene-editing within different cell lines including HEK293, K562 and HeLa. Compared to base editing, PE has a huge advantage in introducing more types of base changes or mutations. Compared to the homologous directed repair (HDR) approach by Cas9, sgRNA and single-stranded oligo (ssODN), gene-edit efficiency of PE was similar but the

percentage of indel mutations was far lower. Therefore, PE is a promising gene-editing method, especially with potential clinical applications, which is required to minimize side effects.

The PLB constructs described in this thesis have been proven to be highly expressed and functional, but there is still space for improvement. Although both linkers $G_6(SGG)_3$ and $(GGGGS)_3$ have been shown to be effective flexible linkers for joining domains requiring a certain degree of interaction or movement [384], linker $G_6(SGG)_3$ have been shown to exhibit a greater ability in enhancing the expression of surface MHC-I molecules containing PLB-S₂₆₉₋₂₇₇ than the $(GGGGS)_3$ linker. The experiments could be repeated using different peptides with the above linkers before confirmation of their contribution to expression and function of PLB molecules. On the other hand, some Gly and Ser residues within the linkers could be replaced by other amino acids such as Ala and Thr to maintain flexibility, or polar amino acids such as Glu and Lys to enhance solubility [385]. Moreover, the copy number “n” of the $(GGGGS)_n$ and $G_6(SGG)_n$ linkers could be adjusted to achieve the optimal length of these GS linkers with respect to fulfilling the optimal distance between the B2M protein and MHC-I binding groove to allow efficient peptide presentation.

The function of PLB molecules was expected to stimulate antigen-specific T-cells and could be evaluated via cytokine production and cytotoxicity. First, enzyme-linked immune absorbent spot (ELISPOT) is a quantitative method for detecting T-cell activation through secreted effector molecules and cytokines, such as IFN- γ , granzyme B and perforin after exposure to specific peptide-MHC complexes [386]. The first step in an ELISPOT assay would be the co-culturing of CD8+ T cell with cells expressing PLB over 24-48 hours followed by capturing cytokines secreted into the

growth medium on plates coated with antibodies to the corresponding cytokine. The amounts of cytokines would indicate whether the cell line expressing the PLB - MHC-I complex could stimulate antigen-specific T-cells. An alternative approach would be to use a traditional cytotoxic T cell assay i.e., the ^{51}Cr release assay, where target cells are labelled with the radioactive chromium and incubated with antigen-specific T cells. Under the cytolytic activity of CTL cells, the ^{51}Cr would be released from the cytoplasm of killed cells into media and quantified.

An obvious next stage in my study would be to determine the spectrum of peptides presented by gene-edited MHC molecules for those expressing the inserted PLB as well as the modified HLA-B27 molecules. For the PLB study, the details of eluted peptides can reveal the stability of the PLB via its effects on peptidomes presented by specific HLA molecules. With respect to the HLA-B27 project, it would be of interest to know the peptidome differences between cell lines expressing HLA-B*27:05, HLA-B*27:05-C67S and HLA-B*27:09. Any significant differences in peptide presentation, especially if novel peptides were determined, would potentially be detrimental in any potential therapeutic application.

Having successfully modified the HLA-B*27:05 cDNA via CRISPR/Cas9 editing, the next step would be to apply these systems to the genomic HLA-B*27:05 sequence. While the designed sgRNA targeting Cys⁶⁷ (sg278F) and Asp¹¹⁶ (sg426F) are still suitable for targeting the desired regions within the HLA-B*27:05 genomic sequence, the donor template introducing the C67S and D116H modifications would require re-designing to include new homology arms containing intron and exon sequences of HLA-B*27:05. An optimal CRISPR/Cas9 system targeting genomic alleles will enable

the future application of this approach with respect to editing HLA-B27 for therapeutic use.

6.3 Final remarks

Thus far, this thesis presents CRISPR/Cas9 technology harnessed to introduce desired modifications into different components of the MHC-I complex. Two independent approaches were developed; (1) generation of stably expressed PLB molecules via the insertion of peptide-linker sequences into the *B2M* locus, and (2) the establishment of genetically modified HLA-B*27:05 molecules expressing C67S, C101S or D116H. In this thesis, three desired peptides and linkers have been successfully knocked-in to exon 1 of the *B2M* locus in HEK cells, with the PLB-S₂₆₉₋₂₇₇-G₆(SGG)₃ protein being expressed and forming MHC-I complexes with HLA class I heavy chain.

To improve the work, several future experiments need to be conducted, including:

1. Optimizing the gene-editing method using different CRISPR/Cas9 platforms, such as Spacer-nick, Primer editor or Base editor, as well as different forms of donor template and delivery such as using viral vectors.
2. Identification of the optimal type and length of linker to maintain the binding between peptide and B2M to ensure the flexibility of peptide for capture by the binding groove of MHC-I molecules.
3. Applying CRISPR-Cas9 to introduce mutations within the genomic HLA-B*27:05 sequence and fully evaluate their biochemical characteristics with respect to monomeric heavy chain folding, ability to aggregate and associate with the MHC-I assembly pathway.
4. Peptide analysis of gene-edited cells to investigate potential changes in the spectrum of peptides presented by MHC-I molecules associating with the wt

B2M and PLB proteins, and between cells expressing HLA-B*27:05 and HLA-B*27 mutations.

5. Evaluate the function of PLB molecules to stimulate antigen-specific T-cells using such assays as the ELISPOT assay to examine cytokine production and ⁵¹Cr release assay to determine cytotoxicity.
6. Generating iPSC cells stably expressing PLB introduced by CRISPR/Cas9 systems and differentiating gene-edited iPSC to iDC which can present desired antigens and provide stimulatory signals for the maturation of T cells.

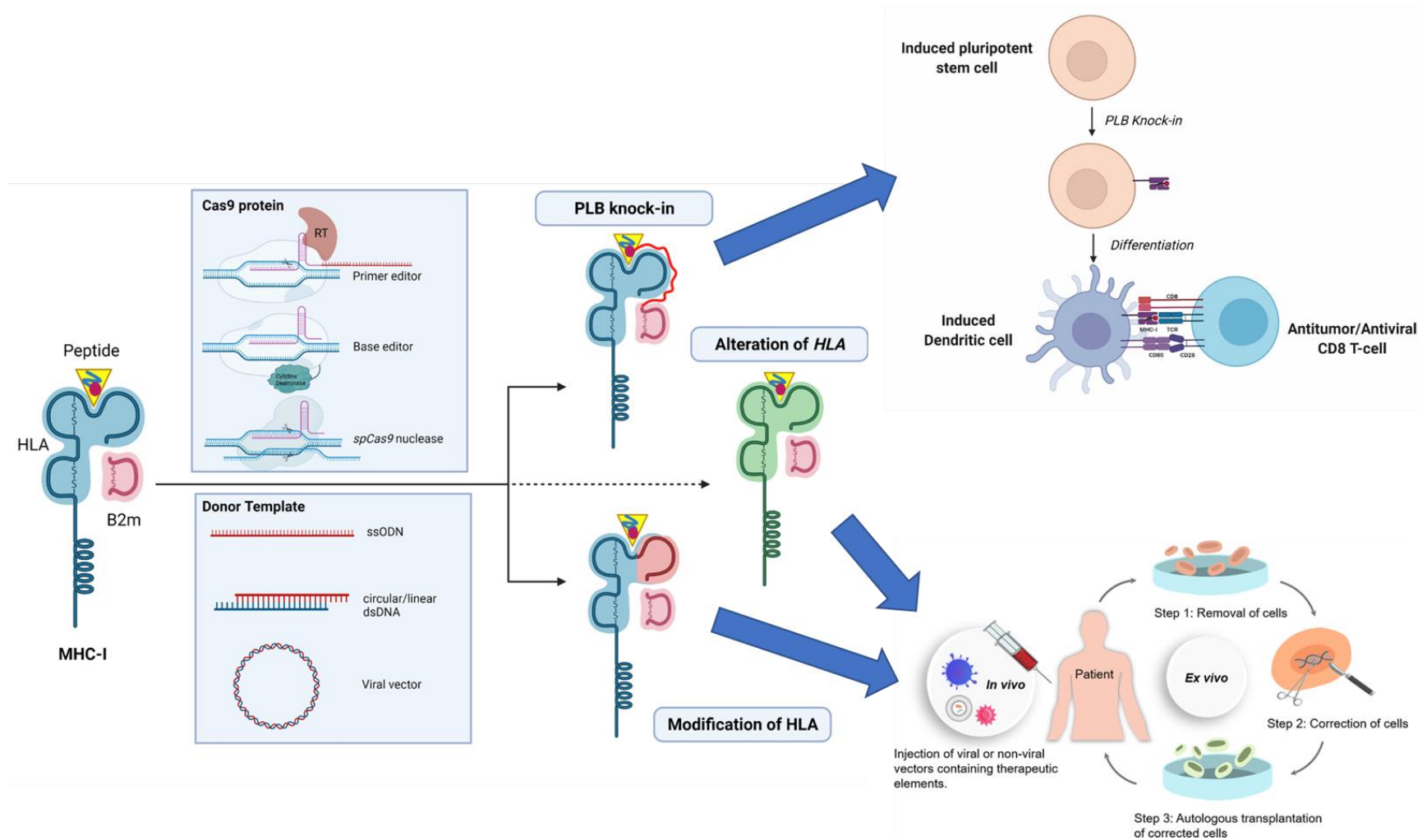


Figure 6.1. CRISPR-mediated modifications of MHC-I complexes and their potential applications.

MHC-I complexes consist of three components: the HLA-I heavy chain, light chain B2m and peptide. Using Cas9 RNP and donor template, a peptide and linker can be fused into the *B2M* loci to produce PLB proteins and nucleotide substitutions can be introduced into the HLA class I heavy chain. CRISPR/Cas9-mediated modifications of MHC-I complexes are promising approaches for developing new immunotherapies and treatments for HLA-related diseases. (i) Induced pluripotent stem cells (iPSC) are potential candidates for PLB knock-in as these cell types can be differentiated to DCs. Gene-edited DCs containing PLB with the desired peptides would be expected to stimulate antigen-specific T cells and direct the immune response against target pathogens or tumors. (ii) Modification of HLA class I heavy chains can be applied *in vivo* or *ex vivo* for treatment of HLA associated diseases. (iii) Gene-editing technology could be exploited to switch between HLA subtypes which are extremely meaningful to overcome current HLA barriers in transplantation.

APPENDIX

Appendix 1: List of chemicals

Product Name	Supplier	Catalogue Number
2-Mercaptoethanol	Sigma-Aldrich	M6250-100ML
Acrylamide	Sigma-Aldrich	01697-500ML
Alt-R CRISPR-Cas9 crRNA	Integrated DNA Technologies	
Alt-R CRISPR-Cas9 tracrRNA	Integrated DNA Technologies	
Alt-R® Cas9 Electroporation Enhancer, 2 nmol	Integrated DNA Technologies	1075915
Alt-R® S.p. Cas9 Nuclease V3, 100 µg	Integrated DNA Technologies	1081058
Amaxa Cell line Nucleofector Kit R	Lonza	VVCA-1001
Amaxa Cell line Nucleofector Kit V	Lonza	VCA-1003
Ammonium persulfate	Merck	7727-54-0
Ampicillin	Sigma-Aldrich	A9393
AMPure XP beads	Beckman	A63880
Anti-GAPDH antibody	Abcam	ab9485
BamHI	New England Biolabs	R0136S
cOmplete™ Protease Inhibitor Cocktail	Roche Applied Science	11697498001
Dithiothreitol (DTT)	Merck	PCG3005
Dimethyl sulfoxide	Merck	200-664-3

dNTP Set 100 mM Solutions	Thermo Scientific™	R0181
EcoRI	New England Biolabs	R0101S
Ethidium bromide solution	Merck	E1510-10ML
Ethylenediaminetetraacetic acid	Sigma-Aldrich	60-00-4
Fetal Bovine Serum	Thermo Scientific™	16000044
Gibco™ DMEM, high glucose, pyruvate	Gibco	11-995-040
HEPES	Merck	H3375-25G
IDTE pH 7.5 (1X TE Solution)	Integrated DNA Technologies	11-01-02-02
IFN-g		
jetPRIME	Polyplus	101000046
LB agar	Merck	L2897
LB broth	Merck	L3022
L-glutamine	Gibco	25030081
Lipofectamine™ CRISPRMAX™ Cas9 Transfection Reagent	Thermo Scientific™	CMAX00003
Magnesium Chloride	Merck	1374248
N,N,N',N'- Tetramethylethylenediamine	Merck	411019-100ML
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs	E2621L

Nextara XT DNA sample preparation kit	Illumina	FC-131-1096
N-methylmaleimide	Merck	389412-5G
NP-40	US Biological	9036-19-5
Nuclease-Free Duplex Buffer	Integrated DNA Technologies	11-01-03-01
Nuclease-Free Water	Merck	W4502-1L
NZYSpeedy Miniprep	Nzytech	MB21001
Opti-MEM™ medium	Thermo Scientific™	31985062
Orange DNA Loading Dye (6X)	Thermo Scientific™	R0631
PBS	Thermo Scientific™	10010023
Penicillin/Streptomycin	Thermo Scientific™	15140122
Phenylmethylsulfonyl fluoride	Merck	52332
Ponceau S	Merck	P3504
Proteinase K	Merck	1.24568
PureLink™ Genomic DNA Mini Kit	Thermo Scientific™	K182001
Q5 High-Fidelity 2X Master Mix	New England Biolabs	M0492S
QIAGEN Plasmid Midi Kit	QIAGEN	12143
QIAquick Gel Extraction Kit	QIAGEN	28706X4
QIAshredder spin column	QIAGEN	79656
QuickExtract™ DNA Extraction Solution	Lucigen	QE09050

QuikChange Site-Directed Mutagenesis Kit	Agilent	200523
RNeasy Mini kit	QIAGEN	74106
RPMI	Thermo Scientific™	11875176
Skimmed milk powder	Great Value	9205097
Sodium chloride	Merck	S9888-500G
SuperSignal™ West Femto Maximum Sensitivity Substrate	Thermo Scientific™	34095
SYBR gold	Thermo Scientific™	S11494
SYBR safe	Thermo Scientific™	S33102
T4 DNA Ligase	Promega	M1801
T7EI nuclease	New England Biolabs	M0302S
TAE (Tris/Acetic Acid/EDTA) Buffer	Bio-rad	1610743
Taq DNA Polymerase with Standard Taq Buffer	New England Biolabs	M0273X
The prestained protein ladder PageRuler	Thermo Scientific™	26616
ThermoScript™ RT	Thermo Scientific™	18090010
Trypan blue	Thermo Scientific™	15250061
Trypsin-EDTA	Thermo Scientific™	25200056

Appendix 2: List of equipment

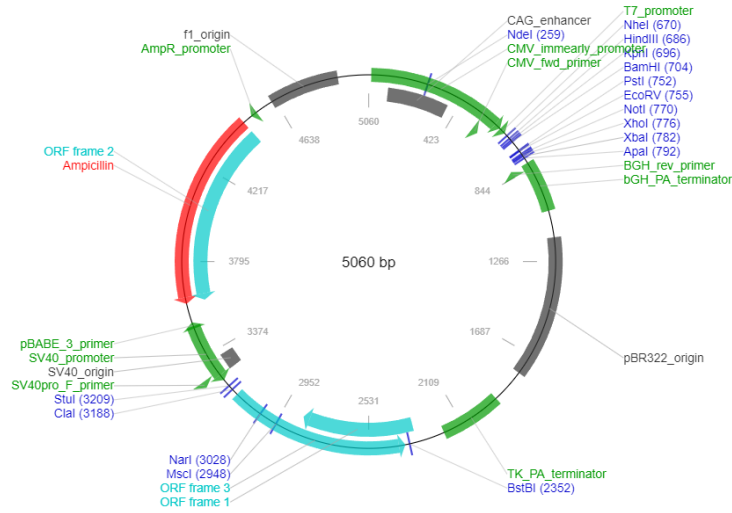
Equipment name	Brand	Model
Bacterial Incubation	Thermo Scientific™	Heratherm™ Compact Microbiological Incubator
Biological Safety Cabinet	Thermo Scientific™	1300 Series A2 Class II Type A2
Cell Centrifuge	Eppendorf	5424R
CO2 Incubators	Thermo Scientific™	Forma™ Series 3 Water Jacketed CO2 Incubator
Cold Centrifuge	Eppendorf	5810R
DNA Electrophoresis Cells	Bio-rad	Mini-Sub Cell GT Systems
Flow Cytometry System	BD	BD FACSCanto™ II Clinical Flow Cytometry System
Gel imaging	Syngene	G:BOX F3
Magnetic stand	Thermo Scientific™	DynaMag™-2 Magnet
NanoDrop One	Thermo Scientific™	ND-ONE-W
Nucleofetor	Lonza	Amaxa Nucleofetor
PCR machine	Bio-rad	C1000 Touch Thermo Cycler
PCR machine	Eppendorf	Mastercycler™ Nexus Thermal Cycler
Precision™ General Purpose Bath	Thermo Scientific™	TSGP02
Protein Electrophoresis Equipment	Bio-rad	Mini-PROTEAN System

Appendix 3: Plasmid map and sequence retrieval

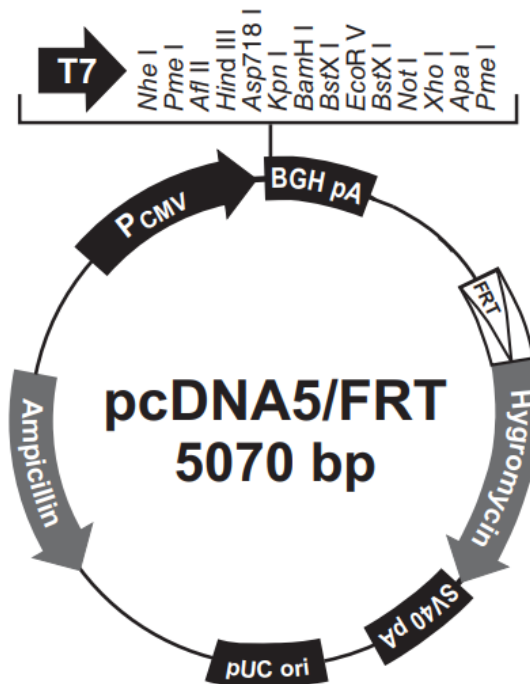
Supplement Data 1. Map of pCR3 and pcDNA™5/FRT Vector.

The figure summarizes the features of the pCR3 vector (A) and pcDNA™5/FRT Vector (B).

A



B



Supplement Data 2. Protein sequence of B2M, HLA-B*27:05, S protein and N protein of SARS-CoV-2 virus.

Name	Protein Sequence
B2M	MSRSVALAVLALLSLSGLEAIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLS FSKDWSFYLLLYTEFTPTTEKDEYACRVNHVTLSPKIVKWDRDM
HLA- B*27:05	MRVTAPRTLLLLLWGAVALTETWAGSHSMRYFHSTVSRPGRGEPFITVGYVDDTLFVRFDSDAASPREEPR APWIEQEGPEYWDRETQICKAKAQTDREDLRTLLRYYNQSEAGSHTLQNMYGCDVGPDRLLRGRYHQDAYD GKDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGECEVWLRRYLENGKETLQRADPPKTHVTH HPISDHEATLRCWALGFYPAEITLTWQRDGEDQTQDELVETRPAGDRFTQKWAAVVPSGEEQRYTCHVQ HEGLPKPLTLRWEPSQSTVPIVGIVAGLAVLAVVVIGAVVAAMCRRKSSGGKGGSSYSQAACSDSAQGS SLTA
S protein	MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTN GTRKRFDNVLPFNDGVYFASTKSNIRGWIFGTTLDSTKQSLIVNNATNVVIVKVEFCQFNDPFLGVYHKN KSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPPQGS LEPLVDLPIGINITRFQTLALHRSYLTGDDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPL ETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNS ASFSTFKCYGVSPKLNLCFTNVYADSFVIRGDEVRIAPGQTKIADYNYKLPDDFTGCVIAWNSNNLDSKV GGNYNYLRLFRKSNLKPFRDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELL HAPATVCGPKKSTNLVKNKCVNFNGLTGTGVLTESNKKFLPFQFGRDIADTTDAVRDPQTLTILEIDITPCSF GGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGNSVFQTRAGCLIGAEHVNNSECDIPI GAGICASYQTQTNPRRARSVASQSIIAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCTMYI CGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS FIEDLLFNKVTLADAGFIKQYGDCLGDIARDLCAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGA ALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDLSSTASALGKLQDVVNQNAQALNTLVKQL SSNFGAISSVLDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVD FCGKGYHLMSFPQSAPHGVVFLHVTVPAQEKNFTTAPAICHGDKAHFPREGVFSNGTHWFVTQRNFYEP QIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLNE VAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCCMTSCCCLKGCCSCGSKCFDEDDSEP VLKGVKLHYT
N protein	MSDNGPQNQRNAPRITFGGSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHGKEDLKFPRGQ GVPINTNSSPDDQIGYYRRATRIRGGDGKMKDLSRWYFYLLGTGPEAGLPYGANKDGIWVATEGALNTPK DHIGTRNPANNAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSSRSRNSTRNTPGSSRGTSPARMAGNG GDAALALLLDRLNQLESKMSGKQQQQQTVTKSAAEASKPRQKRTATKAYNVTQAFGRRGPEQTQGN FGDQELIRQGTDYKHWPIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDKDPNFKDQVILLNKHIDA YKTFPPTPEPKDKKKKADETQALPQRQKQQTVTLPAADLDDFSKQLQQSMSSADSTQA

Supplement Data 3. Amino acid sequences of used PLB proteins

Peptide sequence	Linker sequence	gBlock sequence
TIHDIILECV	(GGGS) ₃	TACTCTCTCTTTCTGGCCTGGAGGCTACTATACATGATATAATA TTAGAATGTGTGGGAGGTGGAGGTTTCAGGAGGTGGAGGATCA GGAGGTGGTGGATCAATCCAGCGTACTCCAAAGATTCAGGTT TAC
TIHDIILECV	G6(SGG) ₃	TACTCTCTCTTTCTGGCCTGGAGGCTACCATCCACGACATCAT CCTGGAGTGCCTGGGAGGTGGAGGTGGAGGTTTCAGGAGGTT CAGGTGGATCAGGAGGTATCCAGCGTACTCCAAAGATTCAGG TTTAC
GMSRIGMEV	(GGGS) ₃	TACTCTCTCTTTCTGGCCTGGAGGCTGGAATGTCGCGCATTG GCATGGAAGTCGGAGGTGGAGGTTTCAGGAGGTGGAGGATCA GGAGGTGGTGGATCAATCCAGCGTACTCCAAAGATTCAGGTT TAC
YLQPRTFLL	(GGGS) ₃	TACTCTCTCTTTCTGGCCTGGAGGCTTATCTTCAACCTAGGAC TTTTCTATTAGGAGGTGGAGGTGGAGGTTTCAGGAGGTTTCAGG TGGATCAGGAGGTATCCAGCGTACTCCAAAGATTCAGGTTTAC

Supplement Data 4. Tables of gBlock sequences containing peptides and linkages

Peptide sequence	Linker sequence	Primers
YLQPRTFLL	G6(SGG)3	<p>Sense-GA.A2/S268-277: TAATAGAAAAGTCCTAGGTTGAAGATAAGCCTCCAGGC CAGAAAGAGAGAGTAGCGCGAG</p> <p>AntiSense-GA.A2/S268-277: TATCTTCAACCTAGGACTTTTCTATTAGGAGGTGGGGG ATCAGGAGGTGGGGGATCAGGA</p>
TIHDIILECV	G6(SGG)3	<p>Sense-GA.A2/HPV.E6 (optimised): CACGCACTCCAGGATGATGTCGTGGATGGTAGCCTCCA GGCCAGAAAGAGAGAGTAGCGC</p> <p>AntiSense-GA.A2/HPV.E6 (optimised): ACCATCCACGACATCATCCTGGAGTGCGTGGGAGGTG GGGGATCAGGAGGTGGGGGATCAGGA</p>

Supplement Data 5. Detail of donor templates used for facilitating HDR of PLB.

Sequences of B2M.exon1, peptides and linkers were highlighted in yellow, brown and blue, respectively.

Name	Type	Length (bp)	Sequence
dsODN.B2M.Silen tM	gBlock	135	gacagcattcgggccgagatgtctcgtccgtggccttagctgtgctcgcgctactctctcttctggctcaga/agctatccagcgtgagtctctctaccctcccctctggtcctctctcccgtctgcac
dsODN.E629-38/(GGGGS)3	gBlock	199	agatgtctcgtccgtggccttagctgtgctcgcgctactctctcttctggcctggaggctactatacatgatataa tattagaatgtgtggagggtgggggatcaggagggtgggggatcaggagggtgggggatcaatccagcgtgagtc tctctaccctcccgtctggtcctctctcccgtctgcacccctct
ssODN.E629-38/(GGGGS)3	Megamer	450	tgcgtcgtcggctggagacaggtgacggctcctgcccgttgctctgattggctggcagcgtttaataataag tggaggcgtcgcgctggcggcattcctgaagctgacagcattcgggccgaatgtctcgtccgtggccttagc tgtcctcgcgctactctctcttctggcctggaggctactatacatgatataaattagaatgtgtggagggtgg gatcaggagggtgggggatcaggagggtgggggatcaatccagcgtgagtcctctaccctcccgtctggtcctt cctctcccgtctgcacccctctgtggccctcgtgtctctcgtcctcgtgacttcccttccaagtctcctgtgt ggcccgcctggggctagctccagggtgatctcggggaagcggcggggctggcctggagtggggaaag
ssODN.E629-38/G6(SGG)3	Megamer	484	tctcttaacctggcactcgtcgtcgtggcctggagacaggtgacggctcctgcccgttgctctgattggctggg cacgcgtttaataataaaggaggcgtcgcctggcgggattcctgaagctgacagcattcgggccgaatgtct cgctccgtggccttagctgtgctcgcgctactctctcttctggcctggaggctaccatccacgacatcatcctgga gtgcgtggagggtggagggtgagggttcaggagggtcagggtggatcaggaggatccagcgtgagtcctctcac cctcccgtctggtcctctctcccgtctgcacccctctgtggccctcgtgtgctctctcgtcctcgtgacttccctt ctccaagtctcctgtggcccgcctggggctagtcagggtggatctcggggaagcggcgggggtggcctgg agtggggaaggggtgcgccccggga
ssODN.N325-333/(GGGGS)3	Megamer	481	tctcttaacctggcactcgtcgtcgtggcctggagacaggtgacggctcctgcccgttgctctgattggctggg cacgcgtttaataataaaggaggcgtcgcctggcgggattcctgaagctgacagcattcgggccgaatgtct cgctccgtggccttagctgtgctcgcgctactctctcttctggcctggaggctggaatgctcgcattggcatgga agtcggagggtgggggatcaggagggtgggggatcaggagggtgggggatcaatccagcgtgagtcctctaccct cccgctctggtcctctctcccgtctgcacccctctgtggccctcgtgtgctctctcgtcctcgtgacttcccttctc caagttctccttggcccgcctggggctagtcagggtggatctcggggaagcggcgggggtggcctggga gtggggaaggggtgcgccccggga
ssODN.S269-277/G6(SGG)3	Megamer	481	tctcttaacctggcactcgtcgtcgtggcctggagacaggtgacggctcctgcccgttgctctgattggctggg cacgcgtttaataataaaggaggcgtcgcctggcgggattcctgaagctgacagcattcgggccgaatgtct cgctccgtggccttagctgtgctcgcgctactctctcttctggcctggaggcttacttcaacctaggactttctat taggaggaggaggaggatcaggaggatcaggaggatcaggaggaaatccagcgtgagtcctctaccctc ccgctctggtcctctctcccgtctgcacccctctgtggccctcgtgtgctctctcgtcctcgtgacttcccttctcc aagtctccttggcccgcctggggctagtcagggtggatctcggggaagcggcgggggtggcctgggag tggggaaggggtgcgccccggga

Appendix 4: Data analysis

Supplement Data 6. Flow-cytometry analysis of MHC-I expression

	Hek	UTC	pCR3	pCR3-B2m	pCR3-A2.HPV-GGGGSx3-B2m	pCR3-A2.Covid.N-GGGGSx3-B2m	pCR3-A2.Covid.S-GGGGSx3-B2m	pCR3-A2.Covid.S-Gx6.SGGx3-B2m
Number of values	3	3	3	3	3	3	3	3
Minimum	97.63	0.47	0.23	70.49	28.43	40.22	38.08	41.97
Maximum	99.64	1.24	0.44	91.9	36.17	41.69	41.74	52.29
Range	2.01	0.77	0.21	21.41	7.74	1.47	3.66	10.32
Mean	98.83	0.7367	0.33	82.51	31.85	41.02	39.86	47.88
Std. Deviation	1.062	0.4362	0.1054	10.94	3.948	0.7427	1.832	5.321
Std. Error of Mean	0.6132	0.2518	0.06083	6.319	2.279	0.4288	1.058	3.072

Number of families	1				
Number of comparisons per family	28				
Alpha	0.05				
Tukey's multiple comparisons test	Mean Diff.	95.00% CI	Below thre	Summary	Adjusted P
Hek vs. UTC	98.1	84.10 to 111.1	Yes	****	<0.0001
Hek vs. pCR3	98.5	84.50 to 112.5	Yes	****	<0.0001
Hek vs. pCR3-B2m	16.32	2.324 to 30.32	Yes	*	0.0175
Hek vs. pCR3-A2.HPV-GGGGSx3-B2m	66.98	52.98 to 80.98	Yes	****	<0.0001
Hek vs. pCR3-A2.Covid.N-GGGGSx3-B2m	57.82	43.82 to 71.82	Yes	****	<0.0001
Hek vs. pCR3-A2.Covid.S-GGGGSx3-B2m	58.97	44.97 to 72.97	Yes	****	<0.0001
Hek vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	50.95	36.95 to 64.95	Yes	****	<0.0001
UTC vs. pCR3	0.4067	-13.59 to 14.72	No	ns	>0.9999
UTC vs. pCR3-B2m	-81.77	-95.77 to -67.77	Yes	****	<0.0001
UTC vs. pCR3-A2.HPV-GGGGSx3-B2m	-31.11	-45.11 to -17.11	Yes	****	<0.0001
UTC vs. pCR3-A2.Covid.N-GGGGSx3-B2m	-40.28	-54.28 to -26.28	Yes	****	<0.0001
UTC vs. pCR3-A2.Covid.S-GGGGSx3-B2m	-39.13	-53.13 to -25.13	Yes	****	<0.0001
UTC vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-47.14	-61.14 to -33.14	Yes	****	<0.0001
pCR3 vs. pCR3-B2m	-82.18	-96.18 to -68.18	Yes	****	<0.0001
pCR3 vs. pCR3-A2.HPV-GGGGSx3-B2m	-31.52	-45.52 to -17.52	Yes	****	<0.0001
pCR3 vs. pCR3-A2.Covid.N-GGGGSx3-B2m	-40.69	-54.69 to -26.69	Yes	****	<0.0001
pCR3 vs. pCR3-A2.Covid.S-GGGGSx3-B2m	-39.53	-53.53 to -25.53	Yes	****	<0.0001
pCR3 vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-47.55	-61.55 to -33.55	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.HPV-GGGGSx3-B2m	50.66	36.66 to 64.66	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.Covid.N-GGGGSx3-B2m	41.49	27.49 to 55.49	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.Covid.S-GGGGSx3-B2m	42.65	28.65 to 56.65	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	34.63	20.63 to 48.63	Yes	****	<0.0001
pCR3-A2.HPV-GGGGSx3-B2m vs. pCR3-A2.Covid.N-GGGGSx3-B2m	-9.167	-23.17 to 4.83	No	ns	0.3517
pCR3-A2.HPV-GGGGSx3-B2m vs. pCR3-A2.Covid.S-GGGGSx3-B2m	-8.013	-22.01 to 5.98	No	ns	0.5029
pCR3-A2.HPV-GGGGSx3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-16.03	-30.03 to -2.03	Yes	*	0.02
pCR3-A2.Covid.N-GGGGSx3-B2m vs. pCR3-A2.Covid.S-GGGGSx3-B2m	1.153	-12.85 to 15.54	No	ns	>0.9999
pCR3-A2.Covid.N-GGGGSx3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-6.863	-20.86 to 7.13	No	ns	0.6705
pCR3-A2.Covid.S-GGGGSx3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-8.017	-22.02 to 5.99	No	ns	0.5025

Supplement Data 7. Flow-cytometry analysis of HLA-A02 expression

	Hek	UTC	pCR3	pCR3 -B2m	pCR3- A2.HPV- GGGSx 3-B2m	pCR3- A2.Covid. N- GGGSx3 -B2m	pCR3- A2.Covid. S- GGGSx3 -B2m	pCR3- A2.Covid.S - Gx6.SGGx 3-B2m
Number of values	3	3	3	3	3	3	3	3
Minimum	93.52	0.34	0.29	63.32	25.33	49.23	42.1	77.87
Maximum	97.69	2.49	0.42	66.61	32.36	50.73	44.38	78.42
Range	4.17	2.15	0.13	3.29	7.03	1.5	2.28	0.55
Mean	96.1	1.21	0.37	65.35	29.35	49.98	43.2	78.12
Std. Deviation	2.252	1.132	0.07	1.773	3.622	0.75	1.142	0.2793
Std. Error of Mean	1.3	0.6536	0.04041	1.024	2.091	0.433	0.6596	0.1613

Compare column means (main column effect)					
Number of families	1				
Number of comparisons per family	28				
Alpha	0.05				
Tukey's multiple comparisons test	Mean Diff.	95.00% CI	Below thre	Summary	Adjusted P
Hek vs. UTC	94.89	89.64 to 100.14	Yes	****	<0.0001
Hek vs. pCR3	95.73	90.48 to 100.98	Yes	****	<0.0001
Hek vs. pCR3-B2m	30.75	25.50 to 36.00	Yes	****	<0.0001
Hek vs. pCR3-A2.HPV-GGGGSx3-B2m	66.75	61.50 to 72.00	Yes	****	<0.0001
Hek vs. pCR3-A2.Covid.N-GGGGSx3-B2m	46.11	40.86 to 51.36	Yes	****	<0.0001
Hek vs. pCR3-A2.Covid.S-GGGGSx3-B2m	52.9	47.65 to 58.15	Yes	****	<0.0001
Hek vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	17.98	12.73 to 23.23	Yes	****	<0.0001
UTC vs. pCR3	0.84	-4.409 to 6.009	No	ns	0.9988
UTC vs. pCR3-B2m	-64.14	-69.39 to -58.89	Yes	****	<0.0001
UTC vs. pCR3-A2.HPV-GGGGSx3-B2m	-28.14	-33.39 to -22.89	Yes	****	<0.0001
UTC vs. pCR3-A2.Covid.N-GGGGSx3-B2m	-48.77	-54.02 to -43.52	Yes	****	<0.0001
UTC vs. pCR3-A2.Covid.S-GGGGSx3-B2m	-41.99	-47.24 to -36.74	Yes	****	<0.0001
UTC vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-76.91	-82.16 to -71.66	Yes	****	<0.0001
pCR3 vs. pCR3-B2m	-64.98	-70.23 to -59.73	Yes	****	<0.0001
pCR3 vs. pCR3-A2.HPV-GGGGSx3-B2m	-28.98	-34.23 to -23.73	Yes	****	<0.0001
pCR3 vs. pCR3-A2.Covid.N-GGGGSx3-B2m	-49.61	-54.86 to -44.36	Yes	****	<0.0001
pCR3 vs. pCR3-A2.Covid.S-GGGGSx3-B2m	-42.83	-48.08 to -37.58	Yes	****	<0.0001
pCR3 vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-77.75	-83.00 to -72.50	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.HPV-GGGGSx3-B2m	36	30.75 to 41.25	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.Covid.N-GGGGSx3-B2m	15.36	10.11 to 20.61	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.Covid.S-GGGGSx3-B2m	22.15	16.90 to 27.40	Yes	****	<0.0001
pCR3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-12.77	-18.02 to -7.52	Yes	****	<0.0001
pCR3-A2.HPV-GGGGSx3-B2m vs. pCR3-A2.Covid.N-GGGGSx3-B2m	-20.63	-25.88 to -15.38	Yes	****	<0.0001
pCR3-A2.HPV-GGGGSx3-B2m vs. pCR3-A2.Covid.S-GGGGSx3-B2m	-13.85	-19.10 to -8.60	Yes	****	<0.0001
pCR3-A2.HPV-GGGGSx3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-48.77	-54.02 to -43.52	Yes	****	<0.0001
pCR3-A2.Covid.N-GGGGSx3-B2m vs. pCR3-A2.Covid.S-GGGGSx3-B2m	6.787	1.538 to 12.036	Yes	**	0.0078
pCR3-A2.Covid.N-GGGGSx3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-28.13	-33.38 to -22.88	Yes	****	<0.0001
pCR3-A2.Covid.S-GGGGSx3-B2m vs. pCR3-A2.Covid.S-Gx6.SGGx3-B2m	-34.92	-40.17 to -29.67	Yes	****	<0.0001

Supplement Data 8. Characterization of surface MHC-I expression on HEK-N₃₂₅₋₃₃₃-(GGGS)₃ cells

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold ?	Summary	Adjusted P Value	A-?		
Hek vs. HekB2mKO	33841	29957 to 37726	Yes	****	<0.0001	B	HekB2mKO	
Hek vs. HekA2-N-D3	9782	5898 to 13667	Yes	****	<0.0001	C	HekA2-N-D3	
Hek vs. HekA2-N-B2	21304	17419 to 25188	Yes	****	<0.0001	D	HekA2-N-B2	
Hek vs. HekA2-N-C6	5749	1864 to 9633	Yes	**	0.0035	E	HekA2-N-C6	
Hek vs. HekA2-N-B4	27665	23781 to 31550	Yes	****	<0.0001	F	HekA2-N-B4	
Hek vs. HekA2-N-A5	23165	19281 to 27050	Yes	****	<0.0001	G	HekA2-N-A5	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Hek vs. HekB2mKO	34243	402	33841	1334	3	3	25.38	14
Hek vs. HekA2-N-D3	34243	24461	9782	1334	3	3	7.335	14
Hek vs. HekA2-N-B2	34243	12940	21304	1334	3	3	15.97	14
Hek vs. HekA2-N-C6	34243	28495	5749	1334	3	3	4.311	14
Hek vs. HekA2-N-B4	34243	6578	27665	1334	3	3	20.74	14
Hek vs. HekA2-N-A5	34243	11078	23165	1334	3	3	17.37	14

Supplement Data 9. Characterization of surface HLA-B7 expression on HEK-N₃₂₅₋₃₃₃-(GGGS)₃ cells

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	below threshold ?	Summary	Adjusted P Value	A-?		
Hek vs. HekB2mKO	6228	5619 to 68	Yes	****	<0.0001	B		HekB2mKO
Hek vs. HekA2-N-D3	2470	1860 to 30	Yes	****	<0.0001	C		HekA2-N-D3
Hek vs. HekA2-N-B2	2706	2097 to 33	Yes	****	<0.0001	D		HekA2-N-B2
Hek vs. HekA2-N-C6	1180	571.1 to 1	Yes	***	0.0002	E		HekA2-N-C6
Hek vs. HekA2-N-B4	5657	5048 to 62	Yes	****	<0.0001	F		HekA2-N-B4
Hek vs. HekA2-N-A5	3820	3210 to 44	Yes	****	<0.0001	G		HekA2-N-A5
Hek vs. HekA2-N-NS	-448.3	-1058 to 1	No	ns	0.2037	H		HekA2-N-NS
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Hek vs. HekB2mKO	6676	447.3	6228	208.4	3	3	29.89	16
Hek vs. HekA2-N-D3	6676	4206	2470	208.4	3	3	11.85	16
Hek vs. HekA2-N-B2	6676	3969	2706	208.4	3	3	12.99	16
Hek vs. HekA2-N-C6	6676	5495	1180	208.4	3	3	5.665	16
Hek vs. HekA2-N-B4	6676	1018	5657	208.4	3	3	27.15	16
Hek vs. HekA2-N-A5	6676	2856	3820	208.4	3	3	18.33	16
Hek vs. HekA2-N-NS	6676	7124	-448.3	208.4	3	3	2.152	16

Supplement Data 10. Characterization of surface HLA-A02 expression on HEK-N₃₂₅₋₃₃₃-(GGGS)₃ cells

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold ?	Summary	Adjusted P Value	A-?		
Hek vs. HekB2mKO	6228	5619 to 6838	Yes	****	<0.0001	B	HekB2mKO	
Hek vs. HekA2-N-D3	2470	1860 to 3079	Yes	****	<0.0001	C	HekA2-N-D3	
Hek vs. HekA2-N-B2	2706	2097 to 3316	Yes	****	<0.0001	D	HekA2-N-B2	
Hek vs. HekA2-N-C6	1180	571.1 to 1790	Yes	***	0.0002	E	HekA2-N-C6	
Hek vs. HekA2-N-B4	5657	5048 to 6267	Yes	****	<0.0001	F	HekA2-N-B4	
Hek vs. HekA2-N-A5	3820	3210 to 4429	Yes	****	<0.0001	G	HekA2-N-A5	
Hek vs. HekA2-N-NS	-448.3	-1058 to 160.9	No	ns	0.2037	H	HekA2-N-NS	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Hek vs. HekB2mKO	6676	447.3	6228	208.4	3	3	29.89	16
Hek vs. HekA2-N-D3	6676	4206	2470	208.4	3	3	11.85	16
Hek vs. HekA2-N-B2	6676	3969	2706	208.4	3	3	12.99	16
Hek vs. HekA2-N-C6	6676	5495	1180	208.4	3	3	5.665	16
Hek vs. HekA2-N-B4	6676	1018	5657	208.4	3	3	27.15	16
Hek vs. HekA2-N-A5	6676	2856	3820	208.4	3	3	18.33	16
Hek vs. HekA2-N-NS	6676	7124	-448.3	208.4	3	3	2.152	16

Supplement Data 11. Flow-cytometry analysis of MHC-I expression by HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold ?	Summary	Adjusted P Value	A-?		
Hek vs. HekB2mKO	11323	9325 to 13320	Yes	****	<0.0001	B	HekB2mKO	
Hek vs. Hek-S269-277-G6(SGG)3-1D3	-6538	-8536 to -4540	Yes	****	<0.0001	C	Hek-S269-277-G6(SGG)3-1D3	
Hek vs. Hek-S269-277-G6(SGG)3-2B1	9311	7314 to 11309	Yes	****	<0.0001	D	Hek-S269-277-G6(SGG)3-2B1	
Hek vs. Hek-S269-277-G6(SGG)3-2C6	-1722	-3719 to 275.9	No	ns	0.0999	E	Hek-S269-277-G6(SGG)3-2C6	
Hek vs. Hek-S269-277-G6(SGG)3-3A5	11329	9331 to 13326	Yes	****	<0.0001	F	Hek-S269-277-G6(SGG)3-3A5	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Hek vs. HekB2mKO	11493	170.3	11323	688.5	3	3	16.45	12
Hek vs. Hek-S269-277-G6(SGG)3-1D3	11493	18031	-6538	688.5	3	3	9.496	12
Hek vs. Hek-S269-277-G6(SGG)3-2B1	11493	2182	9311	688.5	3	3	13.52	12
Hek vs. Hek-S269-277-G6(SGG)3-2C6	11493	13215	-1722	688.5	3	3	2.501	12
Hek vs. Hek-S269-277-G6(SGG)3-3A5	11493	164.3	11329	688.5	3	3	16.45	12

Supplement Data 12. Flow-cytometry analysis of HLA-B7 expression by HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold ?	Summary	Adjusted P Value	A-?		
Hek vs. HekB2mKO	1228	500.3 to 1956	Yes	**	0.0015	B	HekB2mKO	
Hek vs. Hek-S269-277-G6(SGG)3-1D3	-3211	-3939 to -2483	Yes	****	<0.0001	C	Hek-S269-277-G6(SGG)3-1D3	
Hek vs. Hek-S269-277-G6(SGG)3-2B1	453.7	-274.3 to 1182	No	ns	0.3011	D	Hek-S269-277-G6(SGG)3-2B1	
Hek vs. Hek-S269-277-G6(SGG)3-2C6	627	-101.0 to 1355	No	ns	0.1002	E	Hek-S269-277-G6(SGG)3-2C6	
Hek vs. Hek-S269-277-G6(SGG)3-3A5	1276	547.7 to 2004	Yes	**	0.0012	F	Hek-S269-277-G6(SGG)3-3A5	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Hek vs. HekB2mKO	1396	168	1228	250.9	3	3	4.895	12
Hek vs. Hek-S269-277-G6(SGG)3-1D3	1396	4608	-3211	250.9	3	3	12.8	12
Hek vs. Hek-S269-277-G6(SGG)3-2B1	1396	942.7	453.7	250.9	3	3	1.808	12
Hek vs. Hek-S269-277-G6(SGG)3-2C6	1396	769.3	627	250.9	3	3	2.499	12
Hek vs. Hek-S269-277-G6(SGG)3-3A5	1396	120.7	1276	250.9	3	3	5.084	12

Supplement Data 13. Flow-cytometry analysis of HLA-A02 expression by HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold ?	Summary	Adjusted P Value	A-?		
Hek vs. HekB2mKO	4352	3086 to 5618	Yes	****	<0.0001	B	HekB2mKO	
Hek vs. Hek-S269-277-G6(SGG)3-1D3	-9664	-10929 to -8398	Yes	****	<0.0001	C	Hek-S269-277-G6(SGG)3-1D3	
Hek vs. Hek-S269-277-G6(SGG)3-2B1	-468	-1734 to 797.6	No	ns	0.7291	D	Hek-S269-277-G6(SGG)3-2B1	
Hek vs. Hek-S269-277-G6(SGG)3-2C6	-2575	-3841 to -1309	Yes	***	0.0003	E	Hek-S269-277-G6(SGG)3-2C6	
Hek vs. Hek-S269-277-G6(SGG)3-3A5	4391	3125 to 5657	Yes	****	<0.0001	F	Hek-S269-277-G6(SGG)3-3A5	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Hek vs. HekB2mKO	4521	168.7	4352	436.2	3	3	9.977	12
Hek vs. Hek-S269-277-G6(SGG)3-1D3	4521	14184	-9664	436.2	3	3	22.15	12
Hek vs. Hek-S269-277-G6(SGG)3-2B1	4521	4989	-468	436.2	3	3	1.073	12
Hek vs. Hek-S269-277-G6(SGG)3-2C6	4521	7096	-2575	436.2	3	3	5.903	12
Hek vs. Hek-S269-277-G6(SGG)3-3A5	4521	129.7	4391	436.2	3	3	10.07	12

Supplement Data 14. Flow-cytometry analysis of MHC-I expression by HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells.

	K562	K562.B27	K562.B27.C67S	K562.B27.C101S	K562.B27.D116H
Number of values	3	3	3	3	3
Minimum	3671	5668	4545	1829	5681
Maximum	4344	5899	4697	1979	5916
Range	673.0	231.0	152.0	150.0	235.0
Mean	3964	5811	4622	1887	5816
Std. Deviation	344.8	124.9	76.01	80.57	121.5
Std. Error of Mean	199.1	72.13	43.88	46.52	70.14

Number of families	1							
Number of comparisons per family	3							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	B-?		
K562.B27 vs. K562.B27.C67S	1189	946.6 to 1432	Yes	****	<0.0001	C	K562.B27.C67S	
K562.B27 vs. K562.B27.C101S	3924	3681 to 4167	Yes	****	<0.0001	D	K562.B27.C101S	
K562.B27 vs. K562.B27.D116H	-5.333	-248.1 to 237.4	No	ns	0.9999	E	K562.B27.D116H	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
K562.B27 vs. K562.B27.C67S	5811	4622	1189	84.30	3	3	14.11	8
K562.B27 vs. K562.B27.C101S	5811	1887	3924	84.30	3	3	46.55	8
K562.B27 vs. K562.B27.D116H	5811	5816	-5.333	84.30	3	3	0.06327	8

Supplement Data 15. Flow-cytometry analysis of HLA-B27 expression by HEK-S₂₆₉₋₂₇₇-G₆(SGG)₃ cells.

	K562	K562.B27	K562.B27.C67S	K562.B27.C101S	K562.B27.D116H
Number of values	3	3	3	3	3
Minimum	189.0	6672	4366	190.0	6332
Maximum	195.0	7565	4769	300.0	7223
Range	6.000	893.0	403.0	110.0	891.0
Mean	192.7	7262	4555	261.0	6798
Std. Deviation	3.215	511.3	202.7	61.59	447.0
Std. Error of Mean	1.856	295.2	117.0	35.56	258.1

Number of families	1							
Number of comparisons per family	3							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	B-?		
K562.B27 vs. K562.B27.C67S	2708	1871 to 3544	Yes	****	<0.0001	C	K562.B27.C67S	
K562.B27 vs. K562.B27.C101S	7001	6165 to 7838	Yes	****	<0.0001	D	K562.B27.C101S	
K562.B27 vs. K562.B27.D116H	464.0	-372.4 to 1300	No	ns	0.3194	E	K562.B27.D116H	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
K562.B27 vs. K562.B27.C67S	7262	4555	2708	290.4	3	3	9.323	8
K562.B27 vs. K562.B27.C101S	7262	261.0	7001	290.4	3	3	24.11	8
K562.B27 vs. K562.B27.D116H	7262	6798	464.0	290.4	3	3	1.598	8

Appendix 5: Primers for Polymerase Chain Reaction

Experiment	Name	Sequences
PCR	F1.B2M.exon1	AAAATGCAGGTCCGAGCAGT
	R1.B2M.exon1	GACGCTTATCGACGCCCTAA
	F3.B2M.exon2	ATGCAGCGCAATCTCCAGTG
	R3.B2M.exon2	TCTCAGCAGGTGCCACTAATC
	F1.B2McDNA.Cloning	TAAGATGGATCCATGTCTCGCTCCGTGGCCTTA
	R1.B2McDNA.Cloning	AGCCTGGATATCTTACATGTCTCGATCCCCTT
	F-CMV	GATGCGGTTTTGGCAGTACA
	B2MPI-Antisense	AGCCTCCAGGCCAGAAAGAGAGAGTAGCGCGAG CACAGCTAA GGCCACGGAGCGAGACAT
	B2MPI-Sense	ATCCAGCGTACTCCAAGATTCAGGTTTACTCAGC TCATCCAGC AGAGAATGGAAAGTCA
	sense-GA.A2/S ₂₆₈₋₂₇₇	TAATAGAAAAGTCCTAGGTTGAAGATAAGCCTCCA GGCCAGAAAGAGAGAGTAGCGCGAG
	anti-sense- GA.A2/S ₂₆₈₋₂₇₇	TATCTTCAACCTAGGACTTTTCTATTAGGAGGTGG GGGATCAGGAGGTGGGGGATCAGGA
	F1.B2M.cDNA	GCTCGCGTACTCTCTCTTT
	R1.B2M.cDNA	CACGGCAGGCATACTCATCT
	F1-B27	AAGGCACAGACTGACCGAGAG
	R1-B27	GGTGGAGTCCAATCCCAGCAA
	F2-B27	GATGCGGTTTTGGCAGTACA
	R2-B27	CCTCGTTCAGGGCGATGTAA
F4-B27	ATGTCGTAACAACCTCCGCC	

	R4-B27	GATCTCCGCAGGGTAGAAGC
Next Gene Sequencing	F- OT.sgB2M1.3_1	ACTTGGAGAAGGGAAGTCACGG
	R- OT.sgB2M1.3_1	ATAAGTGGAGGCGTCGCGCT
	F- OT.sgB2M1.3_2	GGATAACTGCCTCTTCTCACTGC
	R- OT.sgB2M1.3_2	GTTATAACCCTGAGTTTTCTACCTGCATTATAGTTT C
	F- OT.sgB2M1.3_3	ATGCCTCAGGGAATTGTTGGT
	R- OT.sgB2M1.3_3	ACAAGCAAAAATGGAGGGCTG
	F-OT.sgB27.3_1	GGTCTCACACCCTCCAGAATATG
	R-OT.sgB27.3_1	ACTTGCGCTGGGTGATCTGA
	F-OT.sgB27.3_2	AGCTGGGGAAAGGTGGGC
	R-OT.sgB27.3_2	ATGCAGGGCCCTCCCGCACT
Site Directed-Mutagenesis	F5-SDM.C67S	GTGCCTTGGCCTTAGATATCTGTGTCTCCCGGTC CCAATAC
	R5-SDM.C67S	GTATTGGGACCGGGAGACACAGATATCTAAGGCC AAGGCAC
	F3-SDM.C101S	GGCCCCACGTCGGATCCGTACATATTCTGGAGGG TGTGA
	R3-SDM.C101S	TCACACCCTCCAGAATATGTACGGATCCGACGTG GGGCC
	F4-SDM.D116H	TCCTTGCCGTCGTATGCATGCTGGTGGTACCCGC
	R4-SDM.D116H	GCGGGTACCACCAGCATGCATACGACGGCAAGG A

REFERENCES

REFERENCES

1. Klein, J., *Seeds of time: Fifty years ago Peter A. Gorer discovered the H-2 complex*. Immunogenetics, 1986. **24**(6): p. 331-338.
2. Gorer, P.A., et al., *Studies on the genetic and antigenic basis of tumour transplantation Linkage between a histocompatibility gene and 'fused' in mice*. Proceedings of the Royal Society B: Biological Sciences, 1948. **135**(881): p. 499-505.
3. Carosella, E.D., *Jean Dausset 1916–2009*. Nature Immunology, 2009. **10**(8): p. 797-797.
4. Levine, B.B., A. Ojeda, and B. Benacerraf, *STUDIES ON ARTIFICIAL ANTIGENS. III. THE GENETIC CONTROL OF THE IMMUNE RESPONSE TO HAPTEN-POLY-L-LYSINE CONJUGATES IN GUINEA PIGS*. J Exp Med, 1963. **118**(6): p. 953-7.
5. Trowsdale, J. and J.C. Knight, *Major histocompatibility complex genomics and human disease*. Annu Rev Genomics Hum Genet, 2013. **14**: p. 301-23.
6. Mann, D.L., et al., *Solubilization of Human Leucocyte Membrane Isoantigens*. Nature, 1968. **217**(5134): p. 1180-1181.
7. Robinson, J., et al., *Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles*. PLoS Genet, 2017. **13**(6): p. e1006862.
8. Parham, P., et al., *Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2012. **367**(1590): p. 800-811.
9. Chicz, R.M., et al., *Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size*. Nature, 1992. **358**(6389): p. 764-8.
10. Doyle, C. and J.L. Strominger, *Interaction between CD4 and class II MHC molecules mediates cell adhesion*. Nature, 1987. **330**(6145): p. 256-9.
11. Glazier, K.S., et al., *Germinal center B cells regulate their capability to present antigen by modulation of HLA-DO*. J Exp Med, 2002. **195**(8): p. 1063-9.
12. Leddon, S.A. and A.J. Sant, *Generation of MHC class II-peptide ligands for CD4 T-cell allorecognition of MHC class II molecules*. Curr Opin Organ Transplant, 2010. **15**(4): p. 505-11.
13. Gruen, J.R. and S.M. Weissman, *Human MHC class III and IV genes and disease associations*. Front Biosci, 2001. **6**: p. D960-72.
14. White, P.C., B. Dupont, and M.I. New, *Molecular cloning of steroid 21-hydroxylase*. Endocr Res, 1984. **10**(3-4): p. 335-45.
15. Gur, H., T.D. Geppert, and P.E. Lipsky, *Structural analysis of class I MHC molecules: the cytoplasmic domain is not required for cytoskeletal association, aggregation and internalization*. Mol Immunol, 1997. **34**(2): p. 125-32.
16. Makhdiyeva, D., et al., *MHC class I dimer formation by alteration of the cellular redox environment and induction of apoptosis*. Immunology, 2012. **135**(2): p. 133-9.
17. Park, B., et al., *The Truncated Cytoplasmic Tail of HLA-G Serves a Quality-Control Function in Post-ER Compartments*. Immunity, 2001. **15**(2): p. 213-224.
18. Davis, D.M., et al., *The transmembrane sequence of human histocompatibility leukocyte antigen (HLA)-C as a determinant in inhibition of a subset of natural killer cells*. J Exp Med, 1999. **189**(8): p. 1265-74.
19. Bjorkman, P.J., et al., *Structure of the human class I histocompatibility antigen, HLA-A2*. Nature, 1987. **329**(6139): p. 506-12.
20. Wieczorek, M., et al., *Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation*. Front Immunol, 2017. **8**: p. 292.

21. Norman, P.J., et al., *Co-evolution of human leukocyte antigen (HLA) class I ligands with killer-cell immunoglobulin-like receptors (KIR) in a genetically diverse population of sub-Saharan Africans*. PLoS Genet, 2013. **9**(10): p. e1003938.
22. Petersdorf, E.W., et al., *Major-histocompatibility-complex class I alleles and antigens in hematopoietic-cell transplantation*. N Engl J Med, 2001. **345**(25): p. 1794-800.
23. Fiorillo, M.T., et al., *HLA Class I or Class II and Disease Association: Catch the Difference If You Can*. Frontiers in Immunology, 2017. **8**.
24. Esposito, G., et al., *The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties*. J Mol Biol, 2008. **378**(4): p. 887-97.
25. Drüeke, T.B. and Z.A. Massy, *Beta2-microglobulin*. Semin Dial, 2009. **22**(4): p. 378-80.
26. Ricagno, S., et al., *DE loop mutations affect beta2-microglobulin stability and amyloid aggregation*. Biochem Biophys Res Commun, 2008. **377**(1): p. 146-50.
27. Vincent, M.S., J.E. Gumperz, and M.B. Brenner, *Understanding the function of CD1-restricted T cells*. Nat Immunol, 2003. **4**(6): p. 517-23.
28. Pantopoulos, K., *Function of the hemochromatosis protein HFE: Lessons from animal models*. World J Gastroenterol, 2008. **14**(45): p. 6893-901.
29. Roopenian, D.C. and S. Akilesh, *FcRn: the neonatal Fc receptor comes of age*. Nature Reviews Immunology, 2007. **7**(9): p. 715-725.
30. Rowley, D.R., et al., *Beta-2 microglobulin is mitogenic to PC-3 prostatic carcinoma cells and antagonistic to transforming growth factor beta 1 action*. Cancer Res, 1995. **55**(4): p. 781-6.
31. Relini, A., et al., *Collagen plays an active role in the aggregation of beta2-microglobulin under physiopathological conditions of dialysis-related amyloidosis*. J Biol Chem, 2006. **281**(24): p. 16521-9.
32. Benseny-Cases, N., et al., *Extracellular matrix components modulate different stages in beta2-microglobulin amyloid formation*. J Biol Chem, 2019. **294**(24): p. 9392-9401.
33. Moe, S.M. and S.M. Sprague, *Beta 2-microglobulin induces calcium efflux from cultured neonatal mouse calvariae*. Am J Physiol, 1992. **263**(3 Pt 2): p. F540-5.
34. Rock, K.L., C. Fleischacker, and S. Gamble, *Peptide-priming of cytolytic T cell immunity in vivo using beta 2-microglobulin as an adjuvant*. J Immunol, 1993. **150**(4): p. 1244-52.
35. Mori, M., et al., *Antitumor effect of beta2-microglobulin in leukemic cell-bearing mice via apoptosis-inducing activity: activation of caspase-3 and nuclear factor-kappaB*. Cancer Res, 2001. **61**(11): p. 4414-7.
36. Min, R., et al., *Beta(2)-microglobulin as a negative growth regulator of myeloma cells*. Br J Haematol, 2002. **118**(2): p. 495-505.
37. Gordon, J., et al., *Beta2-microglobulin induces caspase-dependent apoptosis in the CCRF-HSB-2 human leukemia cell line independently of the caspase-3, -8 and -9 pathways but through increased reactive oxygen species*. Int J Cancer, 2003. **103**(3): p. 316-27.
38. Nomura, T., et al., *Beta2-microglobulin promotes the growth of human renal cell carcinoma through the activation of the protein kinase A, cyclic AMP-responsive element-binding protein, and vascular endothelial growth factor axis*. Clin Cancer Res, 2006. **12**(24): p. 7294-305.
39. Huang, W.C., et al., *beta2-microglobulin is a signaling and growth-promoting factor for human prostate cancer bone metastasis*. Cancer Res, 2006. **66**(18): p. 9108-16.
40. Yang, J., et al., *Anti beta2-microglobulin monoclonal antibodies induce apoptosis in myeloma cells by recruiting MHC class I to and excluding growth and survival cytokine receptors from lipid rafts*. Blood, 2007. **110**(8): p. 3028-35.
41. Zhu, Y. and C. Shi, *Beta2-microglobulin, a novel factor for the expansion of mesenchymal stem cells*. Journal of Biotechnology, 2008(136): p. S177.

42. Zhang, B., et al., *Serum Beta-2 Microglobulin: A Possible Biomarker for Atrial Fibrillation*. *Med Sci Monit*, 2021. **27**: p. e932813.
43. Antoniou, A.N., S.J. Powis, and T. Elliott, *Assembly and export of MHC class I peptide ligands*. *Curr Opin Immunol*, 2003. **15**(1): p. 75-81.
44. Johnson, A.E. and M.A. van Waes, *The translocon: a dynamic gateway at the ER membrane*. *Annu Rev Cell Dev Biol*, 1999. **15**: p. 799-842.
45. Ferris, S.P., V.K. Kodali, and R.J. Kaufman, *Glycoprotein folding and quality-control mechanisms in protein-folding diseases*. *Dis Model Mech*, 2014. **7**(3): p. 331-41.
46. Jackson, M.R., et al., *Regulation of MHC class I transport by the molecular chaperone, calnexin (p88, IP90)*. *Science*, 1994. **263**(5145): p. 384-7.
47. Harris, M.R., et al., *Calreticulin and calnexin interact with different protein and glycan determinants during the assembly of MHC class I*. *J Immunol*, 1998. **160**(11): p. 5404-9.
48. Harris, M.R., et al., *Association of ERp57 with mouse MHC class I molecules is tapasin dependent and mimics that of calreticulin and not calnexin*. *J Immunol*, 2001. **166**(11): p. 6686-92.
49. Peaper, D.R. and P. Cresswell, *Regulation of MHC class I assembly and peptide binding*. *Annu Rev Cell Dev Biol*, 2008. **24**: p. 343-68.
50. Springer, S., *Transport and quality control of MHC class I molecules in the early secretory pathway*. *Curr Opin Immunol*, 2015. **34**: p. 83-90.
51. Chapman, D.C. and D.B. Williams, *ER quality control in the biogenesis of MHC class I molecules*. *Semin Cell Dev Biol*, 2010. **21**(5): p. 512-9.
52. Williams, D.B., *Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum*. *J Cell Sci*, 2006. **119**(Pt 4): p. 615-23.
53. Vassilakos, A., et al., *Oligosaccharide binding characteristics of the molecular chaperones calnexin and calreticulin*. *Biochemistry*, 1998. **37**(10): p. 3480-90.
54. Wearsch, P.A., D.R. Peaper, and P. Cresswell, *Essential glycan-dependent interactions optimize MHC class I peptide loading*. *Proceedings of the National Academy of Sciences*, 2011. **108**(12): p. 4950-4955.
55. Park, B., et al., *Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing*. *Cell*, 2006. **127**(2): p. 369-82.
56. Santos, S.G., et al., *Major histocompatibility complex class I-ERp57-tapasin interactions within the peptide-loading complex*. *J Biol Chem*, 2007. **282**(24): p. 17587-93.
57. Schölz, C. and R. Tampé, *The peptide-loading complex--antigen translocation and MHC class I loading*. *Biol Chem*, 2009. **390**(8): p. 783-94.
58. Lewis, J.W. and T. Elliott, *Evidence for successive peptide binding and quality control stages during MHC class I assembly*. *Curr Biol*, 1998. **8**(12): p. 717-20.
59. Wearsch, P.A. and P. Cresswell, *Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer*. *Nat Immunol*, 2007. **8**(8): p. 873-81.
60. Howarth, M., et al., *Tapasin enhances MHC class I peptide presentation according to peptide half-life*. *Proceedings of the National Academy of Sciences*, 2004. **101**(32): p. 11737-11742.
61. Zarling, A.L., et al., *Tapasin is a facilitator, not an editor, of class I MHC peptide binding*. *J Immunol*, 2003. **171**(10): p. 5287-95.
62. Boyle, L.H., et al., *Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway*. *Proc Natl Acad Sci U S A*, 2013. **110**(9): p. 3465-70.
63. Hermann, C., et al., *The binding of TAPBPR and Tapasin to MHC class I is mutually exclusive*. *J Immunol*, 2013. **191**(11): p. 5743-50.

64. Morozov, G.I., et al., *Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing*. Proceedings of the National Academy of Sciences, 2016. **113**(8): p. E1006-E1015.
65. Chang, S.C., et al., *The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism*. Proc Natl Acad Sci U S A, 2005. **102**(47): p. 17107-12.
66. Saveanu, L., et al., *Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum*. Nature Immunology, 2005. **6**(7): p. 689-697.
67. Elliott, T., et al., *Peptide-induced conformational change of the class I heavy chain*. Nature, 1991. **351**(6325): p. 402-6.
68. Kelly, A., et al., *Assembly and function of the two ABC transporter proteins encoded in the human major histocompatibility complex*. Nature, 1992. **355**(6361): p. 641-644.
69. Zacharias, M. and S. Springer, *Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study*. Biophys J, 2004. **87**(4): p. 2203-14.
70. López de Castro, J.A., *How ERAP1 and ERAP2 Shape the Peptidomes of Disease-Associated MHC-I Proteins*. Frontiers in Immunology, 2018. **9**.
71. Smith, K.J., et al., *Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53*. Immunity, 1996. **4**(3): p. 215-28.
72. Walz, S., et al., *The antigenic landscape of multiple myeloma: mass spectrometry (re)defines targets for T-cell-based immunotherapy*. Blood, 2015. **126**(10): p. 1203-13.
73. Rock, K.L., E. Reits, and J. Neefjes, *Present Yourself! By MHC Class I and MHC Class II Molecules*. Trends Immunol, 2016. **37**(11): p. 724-737.
74. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II antigen presentation*. Nature Reviews Immunology, 2011. **11**(12): p. 823-836.
75. van der Velden, V.H. and A.R. Hulsman, *Peptidases: structure, function and modulation of peptide-mediated effects in the human lung*. Clin Exp Allergy, 1999. **29**(4): p. 445-56.
76. Reits, E., et al., *Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I*. Immunity, 2003. **18**(1): p. 97-108.
77. Van Kaer, L., et al., *TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and CD4-8+ T cells*. Cell, 1992. **71**(7): p. 1205-14.
78. Gorbulev, S., R. Abele, and R. Tampé, *Allosteric crosstalk between peptide-binding, transport, and ATP hydrolysis of the ABC transporter TAP*. Proc Natl Acad Sci U S A, 2001. **98**(7): p. 3732-7.
79. Cresswell, P., et al., *The nature of the MHC class I peptide loading complex*. Immunol Rev, 1999. **172**: p. 21-8.
80. Garstka, M.A., et al., *The first step of peptide selection in antigen presentation by MHC class I molecules*. Proceedings of the National Academy of Sciences, 2015. **112**(5): p. 1505-1510.
81. Mpakali, A., et al., *Structural Basis for Antigenic Peptide Recognition and Processing by Endoplasmic Reticulum (ER) Aminopeptidase 2*. J Biol Chem, 2015. **290**(43): p. 26021-32.
82. Hearn, A., I.A. York, and K.L. Rock, *The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum*. J Immunol, 2009. **183**(9): p. 5526-36.
83. Tanioka, T., et al., *Human leukocyte-derived arginine aminopeptidase. The third member of the oxytocinase subfamily of aminopeptidases*. J Biol Chem, 2003. **278**(34): p. 32275-83.

84. Serwold, T., et al., *ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum*. *Nature*, 2002. **419**(6906): p. 480-3.
85. Roelse, J., et al., *Trimming of TAP-translocated peptides in the endoplasmic reticulum and in the cytosol during recycling*. *J Exp Med*, 1994. **180**(5): p. 1591-7.
86. van Hall, T., et al., *The varicellovirus-encoded TAP inhibitor UL49.5 regulates the presentation of CTL epitopes by Qa-1b1*. *J Immunol*, 2007. **178**(2): p. 657-62.
87. Kim, S., et al., *Human cytomegalovirus microRNA miR-US4-1 inhibits CD8(+) T cell responses by targeting the aminopeptidase ERAP1*. *Nat Immunol*, 2011. **12**(10): p. 984-91.
88. Orr, M.T., et al., *Inhibition of MHC class I is a virulence factor in herpes simplex virus infection of mice*. *PLoS pathogens*, 2005. **1**(1): p. e7-e7.
89. Cornel, A.M., I.L. Mimpfen, and S. Nierkens, *MHC Class I Downregulation in Cancer: Underlying Mechanisms and Potential Targets for Cancer Immunotherapy*. *Cancers*, 2020. **12**(7): p. 1760.
90. Maleno, I., et al., *LOH at 6p21.3 region and HLA class I altered phenotypes in bladder carcinomas*. *Immunogenetics*, 2006. **58**(7): p. 503-10.
91. Garrido, M.A., et al., *HLA class I alterations in breast carcinoma are associated with a high frequency of the loss of heterozygosity at chromosomes 6 and 15*. *Immunogenetics*, 2018. **70**(10): p. 647-659.
92. Seliger, B., et al., *Immune escape of melanoma: first evidence of structural alterations in two distinct components of the MHC class I antigen processing pathway*. *Cancer Res*, 2001. **61**(24): p. 8647-50.
93. Lee, J.C., et al., *Elevated TGF-beta1 secretion and down-modulation of NKG2D underlies impaired NK cytotoxicity in cancer patients*. *J Immunol*, 2004. **172**(12): p. 7335-40.
94. Pollard, A.J. and E.M. Bijker, *A guide to vaccinology: from basic principles to new developments*. *Nat Rev Immunol*, 2021. **21**(2): p. 83-100.
95. van Riel, D. and E. de Wit, *Next-generation vaccine platforms for COVID-19*. *Nature Materials*, 2020. **19**(8): p. 810-812.
96. Eisenbarth, S.C., *Dendritic cell subsets in T cell programming: location dictates function*. *Nature Reviews Immunology*, 2019. **19**(2): p. 89-103.
97. Mastelic-Gavillet, B., et al., *Personalized Dendritic Cell Vaccines-Recent Breakthroughs and Encouraging Clinical Results*. *Front Immunol*, 2019. **10**: p. 766.
98. Belderbos, R.A., J. Aerts, and H. Vroman, *Enhancing Dendritic Cell Therapy in Solid Tumors with Immunomodulating Conventional Treatment*. *Mol Ther Oncolytics*, 2019. **13**: p. 67-81.
99. Squadrito, M.L., et al., *EVIR: chimeric receptors that enhance dendritic cell cross-dressing with tumor antigens*. *Nat Methods*, 2018. **15**(3): p. 183-186.
100. Sachamitr, P., et al., *Directed Differentiation of Human Induced Pluripotent Stem Cells into Dendritic Cells Displaying Tolerogenic Properties and Resembling the CD141(+) Subset*. *Front Immunol*, 2017. **8**: p. 1935.
101. Kotsiou, E., J. Brzostek, and K.G. Gould, *Properties and applications of single-chain major histocompatibility complex class I molecules*. *Antioxid Redox Signal*, 2011. **15**(3): p. 645-55.
102. Yu, Y.Y., et al., *Cutting edge: single-chain trimers of MHC class I molecules form stable structures that potently stimulate antigen-specific T cells and B cells*. *J Immunol*, 2002. **168**(7): p. 3145-9.
103. Healy, P.J. and P.S. Helliwell, *Classification of the spondyloarthropathies*. *Curr Opin Rheumatol*, 2005. **17**(4): p. 395-9.
104. Helmick, C.G., et al., *Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I*. *Arthritis Rheum*, 2008. **58**(1): p. 15-25.

105. Shaffrey, C.I. *Surgery for Ankylosing Spondylitis*. Ankylosing Spondylitis 2018 [cited 2018; Available from: <https://www.spineuniverse.com/conditions/spinal-arthritis/ankylosing-spondylitis/surgery-ankylosing-spondylitis>.
106. Sampaio-Barros, P.D., et al., *Primary ankylosing spondylitis: patterns of disease in a Brazilian population of 147 patients*. J Rheumatol, 2001. **28**(3): p. 560-5.
107. Mansour, M., et al., *Ankylosing spondylitis: a contemporary perspective on diagnosis and treatment*. Semin Arthritis Rheum, 2007. **36**(4): p. 210-23.
108. Zhu, W., et al., *Ankylosing spondylitis: etiology, pathogenesis, and treatments*. Bone Res, 2019. **7**: p. 22.
109. Dougados, M., et al., *The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy*. Arthritis Rheum, 1991. **34**(10): p. 1218-27.
110. Brophy, S., et al., *Inflammatory eye, skin, and bowel disease in spondyloarthritis: genetic, phenotypic, and environmental factors*. J Rheumatol, 2001. **28**(12): p. 2667-73.
111. Dean, L.E., et al., *Global prevalence of ankylosing spondylitis*. Rheumatology (Oxford), 2014. **53**(4): p. 650-7.
112. Caffrey, M.F. and D.C. James, *Human lymphocyte antigen association in ankylosing spondylitis*. Nature, 1973. **242**(5393): p. 121.
113. Schlosstein, L., et al., *High association of an HL-A antigen, W27, with ankylosing spondylitis*. N Engl J Med, 1973. **288**(14): p. 704-6.
114. Ellinghaus, D., et al., *Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci*. Nat Genet, 2016. **48**(5): p. 510-8.
115. Taurog, J.D., A. Chhabra, and R.A. Colbert, *Ankylosing Spondylitis and Axial Spondyloarthritis*. New England Journal of Medicine, 2016. **374**(26): p. 2563-2574.
116. Hammer, R.E., et al., *Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human β 2m: An animal model of HLA-B27-associated human disorders*. Cell, 1990. **63**(5): p. 1099-1112.
117. Tran, T.M., et al., *Additional human beta2-microglobulin curbs HLA-B27 misfolding and promotes arthritis and spondylitis without colitis in male HLA-B27-transgenic rats*. Arthritis Rheum, 2006. **54**(4): p. 1317-27.
118. Hammer, G.E., T. Kanaseki, and N. Shastri, *The final touches make perfect the peptide-MHC class I repertoire*. Immunity, 2007. **26**(4): p. 397-406.
119. Reveille, J.D., *An update on the contribution of the MHC to AS susceptibility*. Clin Rheumatol, 2014. **33**(6): p. 749-57.
120. Brown, M.A., et al., *Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment*. Arthritis Rheum, 1997. **40**(10): p. 1823-8.
121. Taurog, J.D., *The mystery of HLA-B27: if it isn't one thing, it's another*. Arthritis Rheum, 2007. **56**(8): p. 2478-81.
122. Khan, M.A., *Polymorphism of HLA-B27: 105 Subtypes Currently Known*. Current Rheumatology Reports, 2013. **15**(10): p. 362.
123. Uchanska-Ziegler, B., A. Ziegler, and P. Schmieder, *Structural and dynamic features of HLA-B27 subtypes*. Curr Opin Rheumatol, 2013. **25**(4): p. 411-8.
124. Wucherpfennig, K.W., *Presentation of a self-peptide in two distinct conformations by a disease-associated HLA-B27 subtype*. J Exp Med, 2004. **199**(2): p. 151-4.
125. Reveille, J.D. and R.M. Maganti, *Subtypes of HLA-B27: history and implications in the pathogenesis of ankylosing spondylitis*. Adv Exp Med Biol, 2009. **649**: p. 159-76.
126. Khan, M.A., *Remarkable polymorphism of HLA-B27: an ongoing saga*. Curr Rheumatol Rep, 2010. **12**(5): p. 337-41.
127. Chatzikyriakidou, A., P.V. Voulgari, and A.A. Drosos, *What is the role of HLA-B27 in spondyloarthropathies?* Autoimmunity Reviews, 2011. **10**(8): p. 464-468.

128. Atagunduz, P., et al., *HLA-B27-restricted CD8+ T cell response to cartilage-derived self peptides in ankylosing spondylitis*. *Arthritis Rheum*, 2005. **52**(3): p. 892-901.
129. Schittenhelm, R.B., et al., *Human Leukocyte Antigen (HLA) B27 Allotype-Specific Binding and Candidate Arthritogenic Peptides Revealed through Heuristic Clustering of Data-independent Acquisition Mass Spectrometry (DIA-MS) Data*. *Mol Cell Proteomics*, 2016. **15**(6): p. 1867-76.
130. Antoniou, A., I. Lenart, and D. Guiliano, *Pathogenicity of Misfolded and Dimeric HLA-B27 Molecules*. *Int J Rheumatol*, 2011. **2011**: p. 486856.
131. Gomez-Simmonds, A. and A.C. Uhlemann, *Clinical Implications of Genomic Adaptation and Evolution of Carbapenem-Resistant *Klebsiella pneumoniae**. *J Infect Dis*, 2017. **215**(suppl_1): p. S18-s27.
132. Ryu, K.H., et al., *Tonsil-derived mesenchymal stromal cells: evaluation of biologic, immunologic and genetic factors for successful banking*. *Cytotherapy*, 2012. **14**(10): p. 1193-202.
133. Albert, L.J. and R.D. Inman, *Molecular mimicry and autoimmunity*. *N Engl J Med*, 1999. **341**(27): p. 2068-74.
134. Fujinami, R.S., et al., *Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease*. *Clin Microbiol Rev*, 2006. **19**(1): p. 80-94.
135. Plot, L. and H. Amital, *Infectious associations of Celiac disease*. *Autoimmun Rev*, 2009. **8**(4): p. 316-9.
136. Ali, F., et al., *Stiff-person syndrome (SPS) and anti-GAD-related CNS degenerations: protean additions to the autoimmune central neuropathies*. *J Autoimmun*, 2011. **37**(2): p. 79-87.
137. Colbert, R.A., T.M. Tran, and G. Layh-Schmitt, *HLA-B27 misfolding and ankylosing spondylitis*. *Mol Immunol*, 2014. **57**(1): p. 44-51.
138. Colbert, R.A., F. Navid, and T. Gill, *The role of HLA-B*27 in spondyloarthritis*. *Best Pract Res Clin Rheumatol*, 2017. **31**(6): p. 797-815.
139. Turner, M.J., et al., *HLA-B27 misfolding in transgenic rats is associated with activation of the unfolded protein response*. *J Immunol*, 2005. **175**(4): p. 2438-48.
140. Mori, K., *Frame switch splicing and regulated intramembrane proteolysis: key words to understand the unfolded protein response*. *Traffic*, 2003. **4**(8): p. 519-28.
141. Wang, P., et al., *The luminal domain of the ER stress sensor protein PERK binds misfolded proteins and thereby triggers PERK oligomerization*. *Journal of Biological Chemistry*, 2018. **293**(11): p. 4110-4121.
142. Yamamoto, K., et al., *Transcriptional induction of mammalian ER quality control proteins is mediated by single or combined action of ATF6alpha and XBP1*. *Dev Cell*, 2007. **13**(3): p. 365-76.
143. Adachi, Y., et al., *ATF6 is a transcription factor specializing in the regulation of quality control proteins in the endoplasmic reticulum*. *Cell Struct Funct*, 2008. **33**(1): p. 75-89.
144. Feng, Y., et al., *Interferon-γ contributes to HLA-B27-associated unfolded protein response in spondyloarthropathies*. *J Rheumatol*, 2012. **39**(3): p. 574-82.
145. Weinreich, S.S., et al., *The role of MHC class I heterodimer expression in mouse ankylosing enthesopathy*. *Immunogenetics*, 1997. **46**(1): p. 35-40.
146. Bettigole, S.E. and L.H. Glimcher, *Endoplasmic reticulum stress in immunity*. *Annu Rev Immunol*, 2015. **33**: p. 107-38.
147. Smith, J.A., *Regulation of Cytokine Production by the Unfolded Protein Response; Implications for Infection and Autoimmunity*. *Front Immunol*, 2018. **9**: p. 422.
148. Bowness, P., *HLA-B27*. *Annu Rev Immunol*, 2015. **33**: p. 29-48.
149. Bowness, P., et al., *Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis*. *J Immunol*, 2011. **186**(4): p. 2672-80.
150. Kollnberger, S., et al., *Cell-surface expression and immune receptor recognition of HLA-B27 homodimers*. *Arthritis Rheum*, 2002. **46**(11): p. 2972-82.

151. Wong-Baeza, I., et al., *KIR3DL2 binds to HLA-B27 dimers and free H chains more strongly than other HLA class I and promotes the expansion of T cells in ankylosing spondylitis*. J Immunol, 2013. **190**(7): p. 3216-24.
152. Zenobia, C. and G. Hajishengallis, *Basic biology and role of interleukin-17 in immunity and inflammation*. Periodontol 2000, 2015. **69**(1): p. 142-59.
153. McHugh, K., et al., *Expression of aberrant HLA-B27 molecules is dependent on B27 dosage and peptide supply*. Ann Rheum Dis, 2014. **73**(4): p. 763-70.
154. Giles, J., et al., *HLA-B27 homodimers and free H chains are stronger ligands for leukocyte Ig-like receptor B2 than classical HLA class I*. J Immunol, 2012. **188**(12): p. 6184-93.
155. Breban, M., et al., *Familial and genetic aspects of spondyloarthritis*. Rheum Dis Clin North Am, 2003. **29**(3): p. 575-94.
156. Yamaguchi, A., et al., *Association of HLA-B39 with HLA-B27-negative ankylosing spondylitis and pauciarticular juvenile rheumatoid arthritis in Japanese patients. Evidence for a role of the peptide-anchoring B pocket*. Arthritis Rheum, 1995. **38**(11): p. 1672-7.
157. Wei, J.C., et al., *HLA-B60 and B61 are strongly associated with ankylosing spondylitis in HLA-B27-negative Taiwan Chinese patients*. Rheumatology (Oxford), 2004. **43**(7): p. 839-42.
158. Brown, M.A., T. Kenna, and B.P. Wordsworth, *Genetics of ankylosing spondylitis--insights into pathogenesis*. Nat Rev Rheumatol, 2016. **12**(2): p. 81-91.
159. Cortes, A., et al., *Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci*. Nature Genetics, 2013. **45**(7): p. 730-738.
160. Kanaseki, T., et al., *ERAAP synergizes with MHC class I molecules to make the final cut in the antigenic peptide precursors in the endoplasmic reticulum*. Immunity, 2006. **25**(5): p. 795-806.
161. Robinson, P.C., et al., *ERAP2 is associated with ankylosing spondylitis in HLA-B27-positive and HLA-B27-negative patients*. Ann Rheum Dis, 2015. **74**(8): p. 1627-9.
162. Dixon, A.L., et al., *A genome-wide association study of global gene expression*. Nat Genet, 2007. **39**(10): p. 1202-7.
163. Cui, X., et al., *Identification of ARTS-1 as a novel TNFR1-binding protein that promotes TNFR1 ectodomain shedding*. J Clin Invest, 2002. **110**(4): p. 515-26.
164. Haroon, N., et al., *Serum cytokine receptors in ankylosing spondylitis: relationship to inflammatory markers and endoplasmic reticulum aminopeptidase polymorphisms*. J Rheumatol, 2010. **37**(9): p. 1907-10.
165. Burton, P.R., et al., *Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants*. Nat Genet, 2007. **39**(11): p. 1329-37.
166. Langrish, C.L., et al., *IL-12 and IL-23: master regulators of innate and adaptive immunity*. Immunol Rev, 2004. **202**: p. 96-105.
167. Tang, C., et al., *Interleukin-23: as a drug target for autoimmune inflammatory diseases*. Immunology, 2012. **135**(2): p. 112-24.
168. Appel, H., et al., *Analysis of IL-17(+) cells in facet joints of patients with spondyloarthritis suggests that the innate immune pathway might be of greater relevance than the Th17-mediated adaptive immune response*. Arthritis Res Ther, 2011. **13**(3): p. R95.
169. McKenzie, B.S., R.A. Kastelein, and D.J. Cua, *Understanding the IL-23-IL-17 immune pathway*. Trends Immunol, 2006. **27**(1): p. 17-23.
170. Mahmoudi, M., et al., *New insights toward the pathogenesis of ankylosing spondylitis; genetic variations and epigenetic modifications*. Mod Rheumatol, 2017. **27**(2): p. 198-209.

171. Baldassano, R.N., et al., *Association of variants of the interleukin-23 receptor gene with susceptibility to pediatric Crohn's disease*. Clin Gastroenterol Hepatol, 2007. **5**(8): p. 972-6.
172. Lim Kam Sian, T.C.C., et al., *Allelic association with ankylosing spondylitis fails to correlate with human leukocyte antigen B27 homodimer formation*. Journal of Biological Chemistry, 2019. **294**(52): p. 20185-20195.
173. Khan, M.A., *Update: the twenty subtypes of HLA-B27*. Curr Opin Rheumatol, 2000. **12**(4): p. 235-8.
174. Akkoc, N. and M.A. Khan, *Etiopathogenic role of HLA-B27 alleles in ankylosing spondylitis*. APLAR Journal of Rheumatology, 2005. **8**(3): p. 146-153.
175. Darke, C., et al., *Immunogenetic study of a new HLA allele, B*2723*. Tissue Antigens, 2002. **60**(5): p. 400-3.
176. Madden, D.R., et al., *The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC*. Cell, 1992. **70**(6): p. 1035-48.
177. Dangoria, N.S., et al., *HLA-B27 misfolding is associated with aberrant intermolecular disulfide bond formation (dimerization) in the endoplasmic reticulum*. J Biol Chem, 2002. **277**(26): p. 23459-68.
178. Fussell, H., et al., *Novel detection of in vivo HLA-B27 conformations correlates with ankylosing spondylitis association*. Arthritis Rheum, 2008. **58**(11): p. 3419-24.
179. Lenart, I., et al., *The MHC Class I Heavy Chain Structurally Conserved Cysteines 101 and 164 Participate in HLA-B27 Dimer Formation*. Antioxidants & Redox Signaling, 2012. **16**(1): p. 33-43.
180. Taurog, J.D., et al., *Inflammatory disease in HLA-B27 transgenic rats*. Immunol Rev, 1999. **169**: p. 209-23.
181. Appel, H., et al., *The solvent-inaccessible Cys67 residue of HLA-B27 contributes to T cell recognition of HLA-B27/peptide complexes*. J Immunol, 2004. **173**(11): p. 6564-73.
182. Huang, C.H., et al., *Cancer immunotherapy using a DNA vaccine encoding a single-chain trimer of MHC class I linked to an HPV-16 E6 immunodominant CTL epitope*. Gene Ther, 2005. **12**(15): p. 1180-6.
183. Crew, M.D., et al., *An HLA-E single chain trimer inhibits human NK cell reactivity towards porcine cells*. Mol Immunol, 2005. **42**(10): p. 1205-14.
184. Cheung, Y.K., et al., *Induction of T-cell response by a DNA vaccine encoding a novel HLA-A*0201 severe acute respiratory syndrome coronavirus epitope*. Vaccine, 2007. **25**(32): p. 6070-7.
185. Huang, B., et al., *Intradermal administration of DNA vaccines combining a strategy to bypass antigen processing with a strategy to prolong dendritic cell survival enhances DNA vaccine potency*. Vaccine, 2007. **25**(45): p. 7824-31.
186. Lilienfeld, B.G., et al., *Transgenic expression of HLA-E single chain trimer protects porcine endothelial cells against human natural killer cell-mediated cytotoxicity*. Xenotransplantation, 2007. **14**(2): p. 126-34.
187. Hung, C.F., et al., *A DNA vaccine encoding a single-chain trimer of HLA-A2 linked to human mesothelin peptide generates anti-tumor effects against human mesothelin-expressing tumors*. Vaccine, 2007. **25**(1): p. 127-35.
188. Palmowski, M.J., et al., *A single-chain H-2Db molecule presenting an influenza virus nucleoprotein epitope shows enhanced ability at stimulating CD8+ T cell responses in vivo*. J Immunol, 2009. **182**(8): p. 4565-71.
189. Kim, S., et al., *Single-chain HLA-A2 MHC trimers that incorporate an immunodominant peptide elicit protective T cell immunity against lethal West Nile virus infection*. J Immunol, 2010. **184**(8): p. 4423-30.

190. Ordaz, M.L., N. Larmonier, and L. Lybarger, *DC-expressed MHC class I single-chain trimer-based vaccines prime cytotoxic T lymphocytes against exogenous but not endogenous antigens*. *Cell Immunol*, 2010. **262**(2): p. 141-9.
191. Kotsiou E, B.J., Lenart I, Antoniou AN, Dyson J, Gould KG, *Dimerization of soluble disulfide trap single-chain major histocompatibility complex class I molecules dependent on peptide binding affinity*. *Antioxidants & Redox Signaling*, 2011. **15**(3): p. 635-644.
192. Matsui, M., et al., *Introduction of a point mutation into an HLA class I single-chain trimer induces enhancement of CTL priming and antitumor immunity*. *Mol Ther Methods Clin Dev*, 2014. **1**: p. 14027.
193. Chen, R., et al., *Anti-metastatic effects of DNA vaccine encoding single-chain trimer composed of MHC I and vascular endothelial growth factor receptor 2 peptide*. *Oncol Rep*, 2015. **33**(5): p. 2269-76.
194. Greten, T.F., et al., *Peptide-beta2-microglobulin-MHC fusion molecules bind antigen-specific T cells and can be used for multivalent MHC-Ig complexes*. *J Immunol Methods*, 2002. **271**(1-2): p. 125-35.
195. Tourdot, S. and K.G. Gould, *Competition between MHC class I alleles for cell surface expression alters CTL responses to influenza A virus*. *J Immunol*, 2002. **169**(10): p. 5615-21.
196. Kim, S., et al., *Licensing of natural killer cells by host major histocompatibility complex class I molecules*. *Nature*, 2005. **436**(7051): p. 709-713.
197. Obermann, S., et al., *Peptide-beta2-microglobulin-major histocompatibility complex expressing cells are potent antigen-presenting cells that can generate specific T cells*. *Immunology*, 2007. **122**(1): p. 90-7.
198. Riedl, S.A.B., et al., *Non-Viral Transfection of Human T Lymphocytes*. *Processes*, 2018. **6**(10): p. 188.
199. Kim, T.K. and J.H. Eberwine, *Mammalian cell transfection: the present and the future*. *Anal Bioanal Chem*, 2010. **397**(8): p. 3173-8.
200. Arber, W., *Restriction Endonucleases*. *Angewandte Chemie International Edition in English*, 1978. **17**(2): p. 73-79.
201. Rothstein, R.J., *One-step gene disruption in yeast*. *Methods Enzymol*, 1983. **101**: p. 202-11.
202. Scherer, S. and R.W. Davis, *Replacement of chromosome segments with altered DNA sequences constructed in vitro*. *Proc Natl Acad Sci U S A*, 1979. **76**(10): p. 4951-5.
203. Smithies, O., et al., *Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination*. *Nature*, 1985. **317**(6034): p. 230-234.
204. Thomas, K.R., K.R. Folger, and M.R. Capecchi, *High frequency targeting of genes to specific sites in the mammalian genome*. *Cell*, 1986. **44**(3): p. 419-28.
205. Mojica, F.J.M., G. Juez, and F. Rodriguez-Valera, *Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites*. *Molecular Microbiology*, 1993. **9**(3): p. 613-621.
206. Klug, A., *The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation*. *Quarterly Reviews of Biophysics*, 2010. **43**(1): p. 1-21.
207. Sternberg, S.H., R.E. Haurwitz, and J.A. Doudna, *Mechanism of substrate selection by a highly specific CRISPR endoribonuclease*. *Rna*, 2012. **18**(4): p. 661-72.
208. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system*. *Nat Protoc*, 2013. **8**(11): p. 2281-2308.
209. Frangoul, H., et al., *CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia*. *N Engl J Med*, 2021. **384**(3): p. 252-260.
210. Scully, R., et al., *DNA double-strand break repair-pathway choice in somatic mammalian cells*. *Nat Rev Mol Cell Biol*, 2019. **20**(11): p. 698-714.

211. Lieber, M.R. and Z.E. Karanjawala, *Ageing, repetitive genomes and DNA damage*. *Nat Rev Mol Cell Biol*, 2004. **5**(1): p. 69-75.
212. Hakem, R., *DNA-damage repair; the good, the bad, and the ugly*. *Embo j*, 2008. **27**(4): p. 589-605.
213. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease*. *Nature*, 2009. **461**(7267): p. 1071-1078.
214. O'Driscoll, M., *Diseases associated with defective responses to DNA damage*. *Cold Spring Harb Perspect Biol*, 2012. **4**(12).
215. Sung, P. and H. Klein, *Mechanism of homologous recombination: mediators and helicases take on regulatory functions*. *Nature Reviews Molecular Cell Biology*, 2006. **7**(10): p. 739-750.
216. Hartlerode, A.J. and R. Scully, *Mechanisms of double-strand break repair in somatic mammalian cells*. *Biochem J*, 2009. **423**(2): p. 157-68.
217. Pannunzio, N.R., G. Watanabe, and M.R. Lieber, *Nonhomologous DNA end-joining for repair of DNA double-strand breaks*. *J Biol Chem*, 2018. **293**(27): p. 10512-10523.
218. Jin, S. and D.T. Weaver, *Double-strand break repair by Ku70 requires heterodimerization with Ku80 and DNA binding functions*. *Embo j*, 1997. **16**(22): p. 6874-85.
219. Spagnolo, L., et al., *Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair*. *Mol Cell*, 2006. **22**(4): p. 511-9.
220. Goodarzi, A.A., et al., *DNA-PK autophosphorylation facilitates Artemis endonuclease activity*. *Embo j*, 2006. **25**(16): p. 3880-9.
221. Ma, Y., K. Schwarz, and M.R. Lieber, *The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps*. *DNA Repair (Amst)*, 2005. **4**(7): p. 845-51.
222. Gottlieb, T.M. and S.P. Jackson, *The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen*. *Cell*, 1993. **72**(1): p. 131-42.
223. Nick McElhinny, S.A., et al., *Ku recruits the XRCC4-ligase IV complex to DNA ends*. *Mol Cell Biol*, 2000. **20**(9): p. 2996-3003.
224. Ahnesorg, P., P. Smith, and S.P. Jackson, *XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining*. *Cell*, 2006. **124**(2): p. 301-13.
225. Buck, D., et al., *Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly*. *Cell*, 2006. **124**(2): p. 287-99.
226. Ochi, T., et al., *DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair*. *Science*, 2015. **347**(6218): p. 185-188.
227. Stinson, B.M., et al., *A Mechanism to Minimize Errors during Non-homologous End Joining*. *Molecular Cell*, 2020. **77**(5): p. 1080-1091.e8.
228. Brandsma, I. and D.C. Gent, *Pathway choice in DNA double strand break repair: observations of a balancing act*. *Genome Integr*, 2012. **3**(1): p. 9.
229. Takata, M., et al., *Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells*. *Embo j*, 1998. **17**(18): p. 5497-508.
230. Lee, J.H. and T.T. Paull, *ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex*. *Science*, 2005. **308**(5721): p. 551-4.
231. Wyman, C. and R. Kanaar, *DNA double-strand break repair: all's well that ends well*. *Annu Rev Genet*, 2006. **40**: p. 363-83.
232. Sugiyama, T., E.M. Zaitseva, and S.C. Kowalczykowski, *A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein*. *J Biol Chem*, 1997. **272**(12): p. 7940-5.
233. Zhao, W., et al., *BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing*. *Nature*, 2017. **550**(7676): p. 360-365.

234. Hicks, W.M., M. Kim, and J.E. Haber, *Increased mutagenesis and unique mutation signature associated with mitotic gene conversion*. *Science*, 2010. **329**(5987): p. 82-5.
235. alstembio. *Custom Genome Editing*. 2022; Available from: <https://www.alstembio.com/web/services/gene-editing.php>.
236. Khalil, A.M., *The genome editing revolution: review*. *J Genet Eng Biotechnol*, 2020. **18**(1): p. 68.
237. Kc, M. and C.J. Steer, *A new era of gene editing for the treatment of human diseases*. *Swiss Med Wkly*, 2019. **149**: p. w20021.
238. Chandrasegaran, S. and D. Carroll, *Origins of Programmable Nucleases for Genome Engineering*. *Journal of Molecular Biology*, 2016. **428**(5, Part B): p. 963-989.
239. Bibikova, M., et al., *Stimulation of homologous recombination through targeted cleavage by chimeric nucleases*. *Mol Cell Biol*, 2001. **21**(1): p. 289-97.
240. Kim, Y.G., J. Cha, and S. Chandrasegaran, *Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain*. *Proc Natl Acad Sci U S A*, 1996. **93**(3): p. 1156-60.
241. Carroll, D., *Genome engineering with zinc-finger nucleases*. *Genetics*, 2011. **188**(4): p. 773-782.
242. Gupta, A., et al., *Targeted chromosomal deletions and inversions in zebrafish*. *Genome Res*, 2013. **23**(6): p. 1008-17.
243. Christian, M.L., et al., *Targeting G with TAL Effectors: A Comparison of Activities of TALENs Constructed with NN and NK Repeat Variable Di-Residues*. *PLOS ONE*, 2012. **7**(9): p. e45383.
244. Römer, P., et al., *Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene*. *Science*, 2007. **318**(5850): p. 645-8.
245. Boch, J., *TALEs of genome targeting*. *Nature Biotechnology*, 2011. **29**(2): p. 135-136.
246. Huo, Z., et al., *Generation of a heterozygous p53 R249S mutant human embryonic stem cell line by TALEN-mediated genome editing*. *Stem Cell Res*, 2019. **34**: p. 101360.
247. Khan, S.H., *Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application*. *Mol Ther Nucleic Acids*, 2019. **16**: p. 326-334.
248. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes*. *Mol Microbiol*, 2002. **43**(6): p. 1565-75.
249. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements*. *J Mol Evol*, 2005. **60**(2): p. 174-82.
250. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity*. *Mol Cell*, 2014. **54**(2): p. 234-44.
251. Bernal-Bernal, D., et al., *Multifactorial control of the expression of a CRISPR-Cas system by an extracytoplasmic function σ /anti- σ pair and a global regulatory complex*. *Nucleic Acids Res*, 2018. **46**(13): p. 6726-6745.
252. Kleinstiver, B.P., et al., *Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing*. *Nature Biotechnology*, 2019. **37**(3): p. 276-282.
253. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems*. *Nature Reviews Microbiology*, 2011. **9**(6): p. 467-477.
254. Jinek, M., et al., *RNA-programmed genome editing in human cells*. *Elife*, 2013. **2**: p. e00471.
255. Gorski, S.A., J. Vogel, and J.A. Doudna, *RNA-based recognition and targeting: sowing the seeds of specificity*. *Nat Rev Mol Cell Biol*, 2017. **18**(4): p. 215-228.
256. Ishino, Y., M. Krupovic, and P. Forterre, *History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology*. *J Bacteriol*, 2018. **200**(7).
257. Sapranauskas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli*. *Nucleic Acids Res*, 2011. **39**(21): p. 9275-82.

258. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. *Science*, 2012. **337**(6096): p. 816-21.
259. Gasiunas, G., et al., *Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. *Proceedings of the National Academy of Sciences*, 2012. **109**(39): p. E2579-E2586.
260. Chiang, T.-W.W., et al., *CRISPR-Cas9D10A nickase-based genotypic and phenotypic screening to enhance genome editing*. *Scientific Reports*, 2016. **6**(1): p. 24356.
261. Komor, A.C., A.H. Badran, and D.R. Liu, *CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes*. *Cell*, 2017. **168**(1-2): p. 20-36.
262. Bakhtiar, A. and E.H. Chowdhury, *PH-responsive strontium nanoparticles for targeted gene therapy against mammary carcinoma cells*. *Asian Journal of Pharmaceutical Sciences*, 2021. **16**(2): p. 236-252.
263. Xu, L., et al., *CRISPR-mediated Genome Editing Restores Dystrophin Expression and Function in mdx Mice*. *Mol Ther*, 2016. **24**(3): p. 564-9.
264. Koo, T., et al., *Selective disruption of an oncogenic mutant allele by CRISPR/Cas9 induces efficient tumor regression*. *Nucleic Acids Res*, 2017. **45**(13): p. 7897-7908.
265. Kleinstiver, B.P., et al., *Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition*. *Nature Biotechnology*, 2015. **33**(12): p. 1293-1298.
266. Villiger, L., et al., *Treatment of a metabolic liver disease by in vivo genome base editing in adult mice*. *Nature Medicine*, 2018. **24**(10): p. 1519-1525.
267. Merienne, N., et al., *The Self-Inactivating KamiCas9 System for the Editing of CNS Disease Genes*. *Cell Rep*, 2017. **20**(12): p. 2980-2991.
268. Lombardo, A., et al., *Site-specific integration and tailoring of cassette design for sustainable gene transfer*. *Nature Methods*, 2011. **8**(10): p. 861-869.
269. Ortinski, P.I., et al., *Integrase-Deficient Lentiviral Vector as an All-in-One Platform for Highly Efficient CRISPR/Cas9-Mediated Gene Editing*. *Molecular therapy. Methods & clinical development*, 2017. **5**: p. 153-164.
270. Cheng, H., F. Zhang, and Y. Ding, *CRISPR/Cas9 Delivery System Engineering for Genome Editing in Therapeutic Applications*. *Pharmaceutics*, 2021. **13**(10).
271. Xu, C.F., et al., *Rational designs of in vivo CRISPR-Cas delivery systems*. *Adv Drug Deliv Rev*, 2021. **168**: p. 3-29.
272. Wang, D., F. Zhang, and G. Gao, *CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors*. *Cell*, 2020. **181**(1): p. 136-150.
273. Wilbie, D., J. Walther, and E. Mastrobattista, *Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing*. *Acc Chem Res*, 2019. **52**(6): p. 1555-1564.
274. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems*. *Science*, 2013. **339**(6121): p. 819-23.
275. Zhang, L., et al., *Lipid nanoparticle-mediated efficient delivery of CRISPR/Cas9 for tumor therapy*. *NPG Asia Materials*, 2017. **9**(10): p. e441-e441.
276. Bai, J., et al., *Engineered targeting tLyp-1 exosomes as gene therapy vectors for efficient delivery of siRNA into lung cancer cells*. *Asian J Pharm Sci*, 2020. **15**(4): p. 461-471.
277. Li, L., et al., *Artificial Virus Delivers CRISPR-Cas9 System for Genome Editing of Cells in Mice*. *ACS Nano*, 2017. **11**(1): p. 95-111.
278. Mali, P., et al., *RNA-guided human genome engineering via Cas9*. *Science*, 2013. **339**(6121): p. 823-6.
279. Finn, J.D., et al., *A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing*. *Cell Rep*, 2018. **22**(9): p. 2227-2235.
280. Farbiak, L., et al., *All-In-One Dendrimer-Based Lipid Nanoparticles Enable Precise HDR-Mediated Gene Editing In Vivo*. *Adv Mater*, 2021. **33**(30): p. e2006619.

281. Usman, W.M., et al., *Efficient RNA drug delivery using red blood cell extracellular vesicles*. Nat Commun, 2018. **9**(1): p. 2359.
282. Lattanzi, A., et al., *Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements*. Mol Ther, 2019. **27**(1): p. 137-150.
283. D'Astolfo, D.S., et al., *Efficient intracellular delivery of native proteins*. Cell, 2015. **161**(3): p. 674-690.
284. German, D.M., et al., *Therapeutic Genome Editing in Cardiovascular Diseases*. JACC Basic Transl Sci, 2019. **4**(1): p. 122-131.
285. Mout, R., et al., *Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing*. ACS Nano, 2017. **11**(3): p. 2452-2458.
286. Hur, J. and A.J. Chung, *Microfluidic and Nanofluidic Intracellular Delivery*. Adv Sci (Weinh), 2021. **8**(15): p. e2004595.
287. Ma, H., et al., *Correction of a pathogenic gene mutation in human embryos*. Nature, 2017. **548**(7668): p. 413-419.
288. Zhang, S., et al., *Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing*. Theranostics, 2021. **11**(2): p. 614-648.
289. Kim, K., et al., *Genome surgery using Cas9 ribonucleoproteins for the treatment of age-related macular degeneration*. Genome Res, 2017. **27**(3): p. 419-426.
290. Savić, N., et al., *In vitro Generation of CRISPR-Cas9 Complexes with Covalently Bound Repair Templates for Genome Editing in Mammalian Cells*. Bio Protoc, 2019. **9**(1).
291. Bak, R.O., et al., *Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6*. Elife, 2017. **6**.
292. Bijlani, S., et al., *The Role of Recombinant AAV in Precise Genome Editing*. Frontiers in Genome Editing, 2022. **3**.
293. Dever, D.P., et al., *CRISPR/Cas9 β -globin gene targeting in human haematopoietic stem cells*. Nature, 2016. **539**(7629): p. 384-389.
294. Yang, H., et al., *One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering*. Cell, 2013. **154**(6): p. 1370-9.
295. Wang, L., et al., *The Application of CRISPR/Cas9 Technology for Cancer Immunotherapy: Current Status and Problems*. Front Oncol, 2021. **11**: p. 704999.
296. Patel, S.J., et al., *Identification of essential genes for cancer immunotherapy*. Nature, 2017. **548**(7669): p. 537-542.
297. Shang, W., et al., *Genome-wide CRISPR screen identifies FAM49B as a key regulator of actin dynamics and T cell activation*. Proc Natl Acad Sci U S A, 2018. **115**(17): p. E4051-e4060.
298. Kaplan, E., et al., *Structure of CYRI-B (FAM49B), a key regulator of cellular actin assembly*. Acta Crystallogr D Struct Biol, 2020. **76**(Pt 10): p. 1015-1024.
299. Hargadon, K.M., C.E. Johnson, and C.J. Williams, *Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors*. Int Immunopharmacol, 2018. **62**: p. 29-39.
300. Xu-Monette, Z.Y., et al., *PD-1/PD-L1 Blockade: Have We Found the Key to Unleash the Antitumor Immune Response?* Front Immunol, 2017. **8**: p. 1597.
301. Shi, L., et al., *CRISPR knock out CTLA-4 enhances the anti-tumor activity of cytotoxic T lymphocytes*. Gene, 2017. **636**: p. 36-41.
302. Zhang, W., et al., *Disruption of CTLA-4 expression on peripheral blood CD8+ T cell enhances anti-tumor efficacy in bladder cancer*. Cancer Chemother Pharmacol, 2019. **83**(5): p. 911-920.
303. Yahata, T., et al., *Programmed cell death ligand 1 disruption by clustered regularly interspaced short palindromic repeats/Cas9-genome editing promotes antitumor immunity and suppresses ovarian cancer progression*. Cancer Sci, 2019. **110**(4): p. 1279-1292.

304. Wang, L., et al., *Chimeric antigen receptor T cell therapy and other therapeutics for malignancies: Combination and opportunity*. *Int Immunopharmacol*, 2019. **70**: p. 498-503.
305. Maude, S.L., et al., *Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia*. *New England Journal of Medicine*, 2014. **371**(16): p. 1507-1517.
306. Ren, J., et al., *Multiplex Genome Editing to Generate Universal CAR T Cells Resistant to PD1 Inhibition*. *Clin Cancer Res*, 2017. **23**(9): p. 2255-2266.
307. Cooper, M.L., et al., *An "off-the-shelf" fratricide-resistant CAR-T for the treatment of T cell hematologic malignancies*. *Leukemia*, 2018. **32**(9): p. 1970-1983.
308. Tang, N., et al., *TGF- β inhibition via CRISPR promotes the long-term efficacy of CAR T cells against solid tumors*. *JCI Insight*, 2020. **5**(4).
309. Gettinger, S., et al., *Impaired HLA Class I Antigen Processing and Presentation as a Mechanism of Acquired Resistance to Immune Checkpoint Inhibitors in Lung Cancer*. *Cancer Discov*, 2017. **7**(12): p. 1420-1435.
310. Hong, C.H., et al., *Antigen Presentation by Individually Transferred HLA Class I Genes in HLA-A, HLA-B, HLA-C Null Human Cell Line Generated Using the Multiplex CRISPR-Cas9 System*. *J Immunother*, 2017. **40**(6): p. 201-210.
311. Mattapally, S., et al., *Human Leukocyte Antigen Class I and II Knockout Human Induced Pluripotent Stem Cell-Derived Cells: Universal Donor for Cell Therapy*. *J Am Heart Assoc*, 2018. **7**(23): p. e010239.
312. Yang, J., et al., *Universal Corneal Epithelial-Like Cells Derived from Human Embryonic Stem Cells for Cellularization of a Corneal Scaffold*. *Transl Vis Sci Technol*, 2018. **7**(5): p. 23.
313. Lorente, E., et al., *Substantial Influence of ERAP2 on the HLA-B*40:02 Peptidome: Implications for HLA-B*27-Negative Ankylosing Spondylitis*. *Mol Cell Proteomics*, 2019. **18**(11): p. 2298-2309.
314. Anzalone, A.V., et al., *Search-and-replace genome editing without double-strand breaks or donor DNA*. *Nature*, 2019. **576**(7785): p. 149-157.
315. Xu, H., et al., *Targeted Disruption of HLA Genes via CRISPR-Cas9 Generates iPSCs with Enhanced Immune Compatibility*. *Cell Stem Cell*, 2019.
316. Zha, S., et al., *Beta-2 microglobulin knockout K562 cell-based artificial antigen presenting cells for ex vivo expansion of T lymphocytes*. *Immunotherapy*, 2019. **11**(11): p. 967-982.
317. Hoerster, K., et al., *HLA Class I Knockout Converts Allogeneic Primary NK Cells Into Suitable Effectors for "Off-the-Shelf" Immunotherapy*. *Front Immunol*, 2020. **11**: p. 586168.
318. Kagoya, Y., et al., *Genetic Ablation of HLA Class I, Class II, and the T-cell Receptor Enables Allogeneic T Cells to Be Used for Adoptive T-cell Therapy*. *Cancer Immunol Res*, 2020. **8**(7): p. 926-936.
319. Kwon, D., et al., *Human Leukocyte Antigen Class I Pseudo-Homozygous Mesenchymal Stem Cells Derived from Human Induced Pluripotent Stem Cells*. *Stem Cell Rev Rep*, 2020. **16**(4): p. 792-808.
320. Petrus-Reurer, S., et al., *Generation of Retinal Pigment Epithelial Cells Derived from Human Embryonic Stem Cells Lacking Human Leukocyte Antigen Class I and II*. *Stem Cell Reports*, 2020. **14**(4): p. 648-662.
321. Shao, L., et al., *Knockout of beta-2 microglobulin enhances cardiac repair by modulating exosome imprinting and inhibiting stem cell-induced immune rejection*. *Cell Mol Life Sci*, 2020. **77**(5): p. 937-952.
322. Zha, S., et al., *Generation of Mesenchymal Stromal Cells with Low Immunogenicity from Human PBMC-Derived beta2 Microglobulin Knockout Induced Pluripotent Stem Cells*. *Cell Transplant*, 2020. **29**: p. 963689720965529.

323. Lamarthee, B., et al., *CRISPR/Cas9-Engineered HLA-Deleted Glomerular Endothelial Cells as a Tool to Predict Pathogenic Non-HLA Antibodies in Kidney Transplant Recipients*. J Am Soc Nephrol, 2021. **32**(12): p. 3231-3251.
324. Lee, O.H., et al., *Generation of a B2M homozygous knockout human somatic cell nuclear transfer-derived embryonic stem cell line using the CRISPR/Cas9 system*. Stem Cell Res, 2021. **59**: p. 102643.
325. Thongsin, N. and M. Wattanapanitch, *CRISPR/Cas9 Ribonucleoprotein Complex-Mediated Efficient B2M Knockout in Human Induced Pluripotent Stem Cells (iPSCs)*. Methods Mol Biol, 2021.
326. Harryvan, T.J., et al., *Enhanced antigen cross-presentation in human colorectal cancer-associated fibroblasts through upregulation of the lysosomal protease cathepsin S*. J Immunother Cancer, 2022. **10**(3).
327. Hirschi, K.K., S. Li, and K. Roy, *Induced pluripotent stem cells for regenerative medicine*. Annu Rev Biomed Eng, 2014. **16**: p. 277-94.
328. Hutchings, S.E. and G.H. Sato, *Growth and maintenance of HeLa cells in serum-free medium supplemented with hormones*. Proc Natl Acad Sci U S A, 1978. **75**(2): p. 901-4.
329. Koefler, H.P. and D.W. Golde, *Human myeloid leukemia cell lines: a review*. Blood, 1980. **56**(3): p. 344-50.
330. Kavsan, V.M., A.V. Iershov, and O.V. Balynska, *Immortalized cells and one oncogene in malignant transformation: old insights on new explanation*. BMC Cell Biology, 2011. **12**(1): p. 23.
331. Luft, T., et al., *Exogenous Peptides Presented by Transporter Associated with Antigen Processing (TAP)-Deficient and TAP-Competent Cells: Intracellular Loading and Kinetics of Presentation*. The Journal of Immunology, 2001. **167**(5): p. 2529-2537.
332. Brodsky, F.M., et al., *Monoclonal antibodies for analysis of the HLA system*. Immunol Rev, 1979. **47**: p. 3-61.
333. Barnstable, C.J., et al., *Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis*. Cell, 1978. **14**(1): p. 9-20.
334. Ellis, S.A., C. Taylor, and A. McMichael, *Recognition of HLA-B27 and related antigen by a monoclonal antibody*. Hum Immunol, 1982. **5**(1): p. 49-59.
335. Stam, N.J., H. Spits, and H.L. Ploegh, *Monoclonal antibodies raised against denatured HLA-B locus heavy chains permit biochemical characterization of certain HLA-C locus products*. J Immunol, 1986. **137**(7): p. 2299-306.
336. O'Gorman, S., D.T. Fox, and G.M. Wahl, *Recombinase-mediated gene activation and site-specific integration in mammalian cells*. Science, 1991. **251**(4999): p. 1351-5.
337. Rodrigues, C., et al., *Allele and haplotype frequencies of HLA-A, B, C, DRB1 and DQB1 genes in polytransfused patients in ethnically diverse populations from Brazil*. Int J Immunogenet, 2015. **42**(5): p. 322-8.
338. Nakagawa, M., et al., *HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein*. J Virol, 2007. **81**(3): p. 1412-23.
339. Kast, W.M., et al., *Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins*. J Immunol, 1994. **152**(8): p. 3904-12.
340. Rensing, M.E., et al., *Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through in vivo and in vitro immunogenicity studies of HLA-A*0201-binding peptides*. J Immunol, 1995. **154**(11): p. 5934-43.
341. Jurtz, V., et al., *NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data*. J Immunol, 2017. **199**(9): p. 3360-3368.

342. Parham, P. and F.M. Brodsky, *Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA-A2 and a variant of HLA-A28*. Hum Immunol, 1981. **3**(4): p. 277-99.
343. Uger, R.A. and B.H. Barber, *Creating CTL targets with epitope-linked beta 2-microglobulin constructs*. J Immunol, 1998. **160**(4): p. 1598-605.
344. Tafuro, S., et al., *Reconstitution of antigen presentation in HLA class I-negative cancer cells with peptide-beta2m fusion molecules*. Eur J Immunol, 2001. **31**(2): p. 440-9.
345. Bicknell, D.C., A. Rowan, and W.F. Bodmer, *Beta 2-microglobulin gene mutations: a study of established colorectal cell lines and fresh tumors*. Proc Natl Acad Sci U S A, 1994. **91**(11): p. 4751-5.
346. Mage, M.G., et al., *A recombinant, soluble, single-chain class I major histocompatibility complex molecule with biological activity*. Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10658-62.
347. Sylvester-Hvid, C., et al., *A single-chain fusion molecule consisting of peptide, major histocompatibility gene complex class I heavy chain and beta2-microglobulin can fold partially correctly, but binds peptide inefficiently*. Scand J Immunol, 1999. **50**(4): p. 355-62.
348. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases*. Nature Methods, 2009. **6**(5): p. 343-345.
349. Uger, R.A., S.M. Chan, and B.H. Barber, *Covalent linkage to beta2-microglobulin enhances the MHC stability and antigenicity of suboptimal CTL epitopes*. J Immunol, 1999. **162**(10): p. 6024-8.
350. McCutcheon, J.A., et al., *HLA-B*0702 antibody epitopes are affected indirectly by distant antigen residues*. Hum Immunol, 1993. **36**(2): p. 69-75.
351. Li, L., M. Dong, and X.G. Wang, *The Implication and Significance of Beta 2 Microglobulin: A Conservative Multifunctional Regulator*. Chin Med J (Engl), 2016. **129**(4): p. 448-55.
352. Cho, S.W., et al., *Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases*. Genome Res, 2014. **24**(1): p. 132-41.
353. Cradick, T.J., et al., *COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites*. Mol Ther Nucleic Acids, 2014. **3**(12): p. e214.
354. Chaurasia, P., et al., *Structural basis of biased T cell receptor recognition of an immunodominant HLA-A2 epitope of the SARS-CoV-2 spike protein*. J Biol Chem, 2021. **297**(3): p. 101065.
355. Montealegre, S., et al., *Dissociation of beta2-microglobulin determines the surface quality control of major histocompatibility complex class I molecules*. FASEB J, 2015. **29**(7): p. 2780-8.
356. Sadeghzadeh, M., et al., *Dendritic cell therapy in cancer treatment; the state-of-the-art*. Life Sci, 2020. **254**: p. 117580.
357. Hilkens, C.M.U. and J.D. Isaacs, *Tolerogenic dendritic cell therapy for rheumatoid arthritis: where are we now?* Clinical & Experimental Immunology, 2013. **172**(2): p. 148-157.
358. Vasaturo, A., et al., *Restoring immunosurveillance by dendritic cell vaccines and manipulation of the tumor microenvironment*. Immunobiology, 2015. **220**(2): p. 243-8.
359. Melief, C.J. and S.H. van der Burg, *Immunotherapy of established (pre)malignant disease by synthetic long peptide vaccines*. Nat Rev Cancer, 2008. **8**(5): p. 351-60.
360. Frankenberger, B. and D.J. Schendel, *Third generation dendritic cell vaccines for tumor immunotherapy*. Eur J Cell Biol, 2012. **91**(1): p. 53-8.
361. Liang, S.-Q., et al., *Genome-wide detection of CRISPR editing in vivo using GUIDE-tag*. Nature Communications, 2022. **13**(1): p. 437.
362. Rana, M.K. and M. Luthra-Guptasarma, *Multi-modal Binding of a 'Self' Peptide by HLA-B*27:04 and B*27:05 Allelic Variants, but not B*27:09 or B*27:06 Variants: Fresh*

- Support for Some Theories Explaining Differential Disease Association. *Protein J*, 2016. **35**(5): p. 346-353.
363. Guiliano, D.B., et al., *Endoplasmic reticulum degradation-enhancing α -mannosidase-like protein 1 targets misfolded HLA-B27 dimers for endoplasmic reticulum-associated degradation*. *Arthritis Rheumatol*, 2014. **66**(11): p. 2976-88.
364. Ploegh, H.L., *Viral strategies of immune evasion*. *Science*, 1998. **280**(5361): p. 248-53.
365. Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9*. *Nat Biotechnol*, 2016. **34**(2): p. 184-191.
366. Liu, M., et al., *Methodologies for Improving HDR Efficiency*. *Front Genet*, 2018. **9**: p. 691.
367. Miura, H., et al., *CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA*. *Sci Rep*, 2015. **5**: p. 12799.
368. Yoshimi, K., et al., *ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes*. *Nat Commun*, 2016. **7**: p. 10431.
369. Antoniou, A.N., et al., *Formation of HLA-B27 homodimers and their relationship to assembly kinetics*. *J Biol Chem*, 2004. **279**(10): p. 8895-902.
370. Tran, T.M., et al., *HLA-B27 in transgenic rats forms disulfide-linked heavy chain oligomers and multimers that bind to the chaperone BiP*. *J Immunol*, 2004. **172**(8): p. 5110-9.
371. Cheng, X., et al., *Molecular mechanism of the susceptibility difference between HLA-B*27:02/04/05 and HLA-B*27:06/09 to ankylosing spondylitis: substitution analysis, MD simulation, QSAR modelling, and in vitro assay*. *SAR QSAR Environ Res*, 2016. **27**(5): p. 409-25.
372. Miyaoka, Y., et al., *Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing*. *Sci Rep*, 2016. **6**: p. 23549.
373. Allen, B., et al., *Non-homologous end joining induced alterations in DNA methylation: A source of permanent epigenetic change*. *Oncotarget*, 2017. **8**(25): p. 40359-40372.
374. Britanova, O.V., et al., *First autologous hematopoietic SCT for ankylosing spondylitis: a case report and clues to understanding the therapy*. *Bone Marrow Transplant*, 2012. **47**(11): p. 1479-81.
375. Booth, A., et al., *Treatment decision-making in sickle cell disease patients*. *J Community Genet*, 2022. **13**(1): p. 143-151.
376. Harrison, C., *First gene therapy for β -thalassemia approved*. *Nat Biotechnol*, 2019. **37**(10): p. 1102-1103.
377. Kim, S., et al., *Single chain MHC I trimer-based DNA vaccines for protection against *Listeria monocytogenes* infection*. *Vaccine*, 2012. **30**(12): p. 2178-86.
378. Ferrari, S., et al., *Gene Editing of Hematopoietic Stem Cells: Hopes and Hurdles Toward Clinical Translation*. *Front Genome Ed*, 2021. **3**: p. 618378.
379. Bak, R.O. and M.H. Porteus, *CRISPR-Mediated Integration of Large Gene Cassettes Using AAV Donor Vectors*. *Cell Rep*, 2017. **20**(3): p. 750-756.
380. Vakulskas, C.A., et al., *A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells*. *Nat Med*, 2018. **24**(8): p. 1216-1224.
381. Dianov, G.L. and U. Hubscher, *Mammalian base excision repair: the forgotten archangel*. *Nucleic Acids Res*, 2013. **41**(6): p. 3483-90.
382. Tran, N.T., et al., *Precise CRISPR-Cas9-mediated gene repair with minimal off-target and unintended on-target mutations in human hematopoietic stem cells*. *Science Advances*, 2022. **8**(22): p. eabm9106.
383. Anzalone, A.V., L.W. Koblan, and D.R. Liu, *Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors*. *Nat Biotechnol*, 2020. **38**(7): p. 824-844.

384. Chen, X., J.L. Zaro, and W.C. Shen, *Fusion protein linkers: property, design and functionality*. *Adv Drug Deliv Rev*, 2013. **65**(10): p. 1357-69.
385. Argos, P., *An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion*. *J Mol Biol*, 1990. **211**(4): p. 943-58.
386. Ranieri, E., G.S. Netti, and M. Gigante, *CTL ELISPOT/ELISPOT Assay and T Cell Detection*, in *Cytotoxic T-Cells: Methods and Protocols*, M. Gigante and E. Ranieri, Editors. 2021, Springer US: New York, NY. p. 65-77.