λ -integrase mediated seamless vector transgenesis platform

Roy, Suki

2020

Roy, Suki. (2020). λ -integrase mediated seamless vector transgenesis platform. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/142767>

<https://doi.org/10.32657/10356/142767>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 21:47:44 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**λ -INTEGRASE MEDIATED SEAMLESS VECTOR
TRANSGENESIS PLATFORM**

SUKI ROY

SCHOOL OF BIOLOGICAL SCIENCES

2020

**λ -INTEGRASE MEDIATED SEAMLESS VECTOR
TRANSGENESIS PLATFORM**

SUKI ROY

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16-01-2020
Date

Suki Roy

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16-01-2020
Date

Associate Professor Peter Dröge

Authorship Attribution Statement

This thesis contains material from one paper published in the following peer-reviewed journal(s) in which I am listed as an author.

Section 3.1 is published as H. Makhija, S. Roy, S. Hoon, F.J. Ghadessy, D. Wong, R. Jaiswal, D. Campana, P. Dröge. (2018). A novel λ integrase-mediated seamless vector transgenesis platform for therapeutic protein expression. Nucleic Acids Research. 46 (16), e99. DOI: 10.1093/nar/gky500.

The contributions of the co-authors are as follows:

- Associate Professor Peter Dröge and Dr Harshyaa Makhija designed the project.
- Associate Professor Peter Dröge, Dr Harshyaa Makhija and I wrote the manuscript.
- Integrase protein purification and seamless vector generation were performed by me.
- Dr Harshyaa Makhija and I performed cell-based targeting assays and analysed the data.
- Dr Shawn Hoon performed sequence and bioinformatic analyses.
- Dr Farid John Ghadessy developed PCR strategies and provided plasmids.
- Dr Dario Campana and Desmond Wong provided the anti-CD19 CAR construct and technical advice.
- Dr Rahul Jaiswal provided advice on Integrase purification and bioinformatics.

16-01-2020
Date

Suki Roy

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to my supervisor, Associate Professor Peter Dröge, for his continuous support, guidance, patience and motivation throughout my PhD. I have been constantly learning from his immense knowledge. His tremendous enthusiasm for science has been a great source of inspiration throughout my research. His tireless guidance helped me at all time. I could not have imagined being supervised and mentored by anyone else for my PhD study.

I will also like to thank my thesis committee members, Associate Professor Oliver Martin Mueller-Cajar and Associate Professor Thirumaran Thanabalu for their insightful comments, encouragement and critical reviews which incited me to widen my research from various perspectives.

I will like to express my deepest gratitude to Dr Harshyaa Makhija for her invaluable contribution in my PhD research and in teaching me all the basics of my PhD project. Her wholehearted support, valuable inputs, critical reviews, constructive suggestions and willingness to help any time without hesitation, requires special mention. I am grateful to Dr Shawn Hoon, Dr Farid John, Dr Dario Campana, Dr Rahul Jaiswal and Desmond Wong for their vital guidance and contribution in my PhD research.

I will like to thank my fellow lab mates, Dr Roland Ivanyi-Nagy, Dr Ramani Priya Dharshana, Sabrina Peter, Syed Moiz Ahmed and Dr Bao Qiuye for providing a joyful and enjoyable working environment during the course of my PhD. I will like to convey special thanks to Dr Namarata Pradeep Chaudhari for her assistance in cell culture work and guiding me in the protein activity and cell viability assay.

I am blessed to have a bunch of genuine friends in my life who form an integral part of my PhD journey. I express my heartfelt gratitude to Ankur Sharma for taking out time to proofread my thesis, helping me to analyse FACS data in FlowJo software and to chip in his meticulous inputs with continuous support throughout my PhD journey. I am eternally grateful to Abhinay Sandupatla and Prativa Das for lending undying support and care over the years and for being a constant source of entertainment when it was needed the most. A special thanks to Vaisak Parekatt for his constant motivation and support in my PhD.

This acknowledgement will remain incomplete without paying special regards to my parents, Mr Subash Chandra Roy and Mrs Rani Roy, who form the center of my existence and the pillar of strength throughout my PhD journey. This thesis would never have seen the light of the day without the two people; my sister, Mrs Sanchita Roy and brother-in-law, Mr Sayantan

Mukherjee. I wish to thank them for proofreading my thesis and their constant support and care all the time.

My heartfelt thanks to Nanyang Technological University for the generous scholarship support throughout the four years and for providing excellent learning and research facilities. Above all, I thank the Almighty for His kind blessings and for instilling in me the motivation to strive harder each day with equal zeal and enthusiasm.

*This thesis is dedicated with love and
gratitude to my beloved parents*

TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	4
LIST OF FIGURES	7
LIST OF TABLES.....	11
ABSTRACT.....	12
1. INTRODUCTION.....	14
1.1. Random Integration Tools.....	14
1.2. Site-Specific Integration Tool	16
1.2.1. Meganucleases.....	18
1.2.2. Zinc finger nucleases (ZFNs)	18
1.2.3. Transcription activator-like effector-based nucleases (TALEN)	19
1.2.4. Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system.....	19
1.2.5. CRISPR-associated Transposase (CAST).....	26
1.2.6. Site-specific integration by recombinases	26
1.2.6.1. Types of serine recombinases.....	28
1.2.6.2. Types of tyrosine recombinases	30
1.2.6.3. λ -Int-mediated site-specific recombination	31
1.3. Seamless vector and its significance	35
1.3.1. Generation of the seamless vector by recombinases	36
1.4. Overview of biologics production platform from the mammalian expression system	36
1.5. Enhancement of transgene expression by introducing a DNA element in the seamless vector..38	
1.5.1. Different DNA elements to enhance transgene expression inside the genome.....	39
1.6. Aim of the project.....	40
2. MATERIALS AND METHODS.....	42
2.1. Cell lines	42
2.2. Plasmids.....	42
2.3. Plasmid DNA purification.....	53
2.4. DNA and RNA quantification	53
2.5. Cell culture	54
2.6. Transfections	54
2.7. Antibiotic Selection and screening for targeted cell clones	55
2.8. Identification of recombination events by PCR screening	55
2.9. Inverse PCR.....	56
2.10. Purification of Int-h/218.....	56
2.11. <i>In vitro</i> recombination assays and seamless vector production.....	57
2.12. Quantification of gel image using ImageJ software	57
2.13. Transgene insertion Mapping	58
2.14. Cell surface staining of anti-CD19 CAR on hESCs.....	59

2.15. Flow cytometry	60
2.16. Southern Blot Analysis	60
2.17. Western blotting	60
2.18. Factor VIII activity Test	61
2.19. MTT assay	62
2.20. <i>In vivo</i> generation of the seamless vector	62
3. RESULTS	63
3.1. λ -Integrase mediated seamless vector transgenesis platform for therapeutic protein expression	63
3.1.1. Strategy to generate seamless DNA target vectors via <i>in vitro</i> intramolecular recombination for subsequent genome editing.....	63
3.1.2. Partial purification of λ Integrase and generation of seamless vector.....	65
3.1.3. <i>In vitro</i> recombination by Int-h/218 with linear substrate plasmid	69
3.1.4. <i>In vitro</i> recombination by Int-h/218 with <i>attH4X</i> x <i>attP4X</i> vs <i>attL4X</i> x <i>attR4X</i>	70
3.1.5. λ -Integrase mediated <i>in vivo</i> production of the seamless vector in <i>E. coli</i> strain LZ41	71
3.1.5.1. Strategy for the generation of the seamless vector <i>in vivo</i>	71
3.1.5.2. Optimization of culture condition of <i>E. coli</i> strain LZ41 for <i>in vivo</i> production of seamless vector	73
3.1.5.3. Large scale <i>in vivo</i> production of the seamless vector from <i>E. coli</i> strain LZ41	77
3.1.6. Site-specific seamless vector insertion into human <i>LINE-1</i>	78
3.1.6.1. Seamless vector transgenesis in HT1080 cells.....	78
3.1.6.2. <i>In vivo</i> sequential intra- and inter-molecular recombination leading to transgenesis of seamless vectors	87
3.1.6.3. Seamless vector transgenesis in human embryonic stem cells (hESCs).....	93
3.1.7. Application of seamless vector in cancer gene therapy	96
3.1.8. Anti-CD19 CAR seamless vector insertion and sustained expression in hESCs.....	98
3.1.8.1. Western blot analysis of anti-CD19 CAR expressing clone #19	103
3.1.8.2. Southern blot analysis of anti-CD19 CAR expressing clone #19	104
3.1.8.3. Long term and stable anti-CD19 CAR expression in clone #19	105
3.1.9. Identification of <i>LINE-1</i> as preferred target loci	106
3.2. λ -Integrase mediated seamless vector transgenesis platform for biologics production	108
3.2.1. Strategy for biologics production platform in HEK293T	108
3.2.2. Targeting of HEK docking clones with seamless vector <i>attL4X-Puro-EF-GFP</i>	110
3.2.3. FACS analysis for EGFP expression of targeted HEK docking clones	113
3.2.4. Southern Blot analysis of parental and targeted HEK docking clones.....	113
3.2.5. Application of seamless vector transgenesis platform for factor VIII protein production.	116
3.2.6. Strategy for FVIII protein production platform in HEK293T.....	118
3.2.7. PCR Screening for HEK docking bulk cultures targeted with <i>pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X</i>	118

3.2.8. FACS sorting and FVIII protein activity test for HEK docking bulk cultures targeted with <i>pattP4X-Puro-EF-F8-IRES-ZsGreen-attH4X</i>	119
3.2.9. Attempts to obtain clones for HEK docking clones targeted with <i>attL4X-Puro-EF-F8-IRES-ZsGreen</i>	122
3.3. Enhancement of transgene expression from site-specifically targeted seamless vector carrying CTCF binding motifs.....	124
3.3.1. CCCTC-binding factor (CTCF)	124
3.3.2. The hypothesis of inclusion of the CTCF sites in the seamless vector for enhanced transgene expression.....	127
3.3.3. Strategy to include CTCF site for enhanced transgene expression	127
3.3.4. FACS analysis for <i>mNeon</i> expression in targeted hES clones	128
3.3.5. PCR analysis for seamless vector insertion at <i>LINE-1 elements</i>	131
3.4. λ -Integrase mediated seamless vector transgenesis in actively transcribed <i>LINE-1 elements</i>	135
3.4.1. Actively transcribed human <i>LINE-1 elements</i>	135
3.4.2. Strategy for seamless vector transgenesis in actively transcribed <i>LINE-1</i>	137
3.4.3. PCR screening for hES clones targeted with <i>attL4X-GFP-EF-Puro-IRES</i>	138
3.4.4. Transgene insertion site mapping to identify the genomic location of <i>attL4X-GFP-EF-Puro-IRES</i> transgenesis in hESCs	139
4. DISCUSSION	141
4.1. λ -Integrase mediated seamless vector transgenesis platform for therapeutic protein expression.....	141
4.2. λ -Integrase mediated seamless vector transgenesis platform for biologics production.....	150
4.3. Enhancement of transgene expression from seamless vector carrying CTCF binding motifs.....	154
4.4. λ -Integrase mediated seamless vector transgenesis in actively transcribed <i>LINE-1 elements</i>	155
5. CONCLUSION	157
6. FUTURE WORK.....	158
REFERENCES.....	159
APPENDIX.....	177
PUBLICATIONS.....	187
CONFERENCE PRESENTATIONS	188

LIST OF ABBREVIATIONS

AAV	Adeno-associated Virus
APC	Antigen Presenting Cells
<i>attB</i>	Bacteria Attachment Site
<i>attP</i>	Phage Attachment Site
BHK	Baby Hamster Kidney
Bpa	Bovine Poly (A)
CAR	Chimeric Antigen Receptor
CCC	Covalently Closed Circular
CHO	Chinese Hamster Ovary cells
Cre	Cyclization recombinase
CRISPR/Cpf1	Clustered Regularly Interspaced Short Palindromic Repeats from <i>Prevotella</i> and <i>Francisella</i> 1
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR associated protein 9
CTCF	CCCTC-binding factor
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic Acid
DSBs	Double-Strand Breaks
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia Coli</i>
EDTA	Ethylenediaminetetraacetic Acid
EF	Elongation Factor
EGFP	Enhanced Green Fluorescent Protein
EGTA	Ethylene Glycol Tetraacetic Acid
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
FIS	Factor for Inversion Stimulation
FLF8	Full-Length Factor 8
FLP	Flippase
FRT	Flippase Recognition Target
FSC	Forward Scatter
FVIII	Factor VIII
sgRNA	Single guide RNA
HCl	Hydrochloric Acid

HDR	Homology-Directed Repair
HEK293-T	Human Embryonic Kidney cells 293-T
hESCs	Human Embryonic Stem Cells
HLA	Human Leukocyte Antigen
HSV	Herpes Simplex Virus
Hygro	Hygromycin
IHF	Integration Host Factor
Int	Integrase
iPSCs	Induced Pluripotent Stem Cells
IRES	Internal Ribosome Entry Site
ITRs	Inverted Terminal Repeats
KCl	Potassium Chloride
LB	Luria-Bertani liquid medium
LCR	Locus Control Regions
<i>LINE-1</i>	<i>Long Interspersed Elements-1</i>
loxP	Locus of X-over P1
LV	Lentiviral Vectors
MC	Minicircle
MP	Miniplasmid
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
Neo	Neomycin
NHEJ	Nonhomologous End Joining
NK	Natural Killer Cells
NTA	Nitrilotriacetic Acid
OC-SV	Open Circular Seamless Vector
ORF	Open Reading Frame
PAM	Protospacer-Associated Motif
PB	Piggy- Bac
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PGK	Phosphoglycerate Kinase
PMSF	Phenylmethylsulfonyl Fluoride
Puro	Puromycin

RDF	Recombination Directionality Factor
RIPA buffer	Radioimmunoprecipitation assay buffer
RMCE	Recombinase Mediated Cassette Exchange
RV	Retroviral
RT	Reverse Transcriptase
S/MARs	Scaffold/Matrix Attachment Regions
SB	Sleeping Beauty
scFv	Single-Chain Variable Fragment
scIHF2	Single Polypeptide Chain IHF
SC-SV	Supercoiled Seamless Vector
SDS	Sodium Dodecyl Sulfate
SSC	Side Scatter
SSR	Site Specific Recombination
TAE	Tris-Acetate-Boric acid
TALEN	Transcription Activator Like Effector Nucleases
TBE	Tris-Boric Acid-EDTA
TCR	T Cell Receptor
TE	Tris-EDTA
TH1	T-helper type 1
TLR9	Toll-Like-Receptor 9
UCOE	Ubiquitous Chromatin Region Opening Element
UTF1	Undifferentiated Embryonic Cell Transcription Factor 1.
UTR	Untranslated Region
VH	Variable Heavy
VL	Variable Light
WT	Wild Type
Xis	Excisionase
ZFN	Zinc Finger Nucleases
ZFP	Zinc Finger Protein
λ -Int	Lambda Integrase

LIST OF FIGURES

Figure 1.1. Different outcomes of genome editing by NHEJ and HDR.	17
Figure 1.2. Schematic diagram of the ZFN system.....	19
Figure 1.3. Schematic diagram of the TALEN system.....	19
Figure 1.4. Schematic representation of the CRISPR/Cas9 system.....	20
Figure 1.5. Different variants of CRISPR/Cas9 tools:.....	24
Figure 1.6. Different applications of CRISPR/Cas9 technology beyond genome engineering.....	25
Figure 1.7. CRISPR-associated transposase (CAST)	26
Figure 1.8. Schematic representation of site-specific recombination by serine and tyrosine recombinase..	28
Figure 1.9. BxB1 integrase mediated integration and excision reaction.	29
Figure 1.10. Nucleotide sequences of the <i>attP</i> and <i>attB</i> sites recognised by phiC31 integrase	29
Figure 1.11. Schematic representation of site-specific recombination by phiC31 integrase.....	30
Figure 1.12. Nucleotide sequences of 34bp <i>loxP</i> site and 48bp <i>FRT</i> site	31
Figure 1.13. Cre recombinase mediated recombination	31
Figure 1.14. λ -Int mediated site-specific recombination.	32
Figure 1.15. Site-specific recombination mediated by λ -Int via a Holliday junction model.....	33
Figure 1.16. A schematic diagram showing λ -Int-mediated recombination between genomic <i>attH4X</i> (located in <i>LINE-1 elements</i>) and <i>attP4X</i> (present on a plasmid target vector).	34
Figure 1.17. Intramolecular recombination in the parental plasmid (PP).....	36
Figure 1.18. Recombinase-mediated cassette exchange (RMCE).....	38
 Figure 2.1. Vector map of <i>pattP4X-PGKssEGFP-bpa-attH4X</i>	43
Figure 2.2. Vector map of <i>pattP4X-PGKss-Puro-bpa-attH4X</i>	43
Figure 2.3. Vector map of <i>pattP4X-PGKss-CAR-bpa-attH4X</i>	43
Figure 2.4. Vector map of <i>pEF1-CAR-bpa</i>	44
Figure 2.5. Vector map of <i>pattP4X-PGKss-Puro-bpa-EF-CAR-bpa-attH4X</i>	44
Figure 2.6. Vector map of <i>pattP4X-PGKss-CAR-bpa-EF-GFP-bpa-attH4X</i>	45
Figure 2.7. Vector map of <i>pattP4X-PGKss-CAR-bpa-EF-GFP-bpa-attH4X</i>	46
Figure 2.8. Vector map of <i>pattP4X-CTCF-UTF1-GFP-CTCF-PGKss-Puro-bpa-attH4X</i>	46
Figure 2.9. Vector map of <i>pattP4X-scCTCF-UTF1-GFP-scCTCF-PGKssPuro-bpa-attH4X</i>	47
Figure 2.10. Vector map of <i>pattP4X-CTCF-UTF1-GFP-PGKss-Puro-bpa-CTCF-attH4X</i>	48
Figure 2.11. Vector maps of <i>pattP4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro-bpa-attH4X</i> , <i>pattP4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro- bpa-attH4X</i> and <i>pattP4X-CTCF-UTF1-mNeon-PGKssPuro- bpa-CTCF-attH4X</i>	48

Figure 2.12. Vector map of <i>pPGKss-4x-attH4X-Neo-bpa</i>	49
Figure 2.13. Vector map of <i>pPGKss-4x-attH4X-Hygro-bpa</i>	49
Figure 2.14. Vector map of <i>pattP4X-Puro-bpa-EF-GFP-bpa-attH4X</i>	50
Figure 2.15. Vector map of <i>pattP4X-Puro-bpa-EF-FLF8-IRES-ZsGreen-polyA-attH4X</i>	51
Figure 2.16. Vector map of <i>pattP4X-bpa-GFP-EF-bpa-Puro-IRES-attH4X</i>	51
Figure 2.17. An illustration depicting the PCR strategy for mapping insertion sites of seamless vector insertion in HT1080 cells and hESCs.	59
Figure 3.1. λ -Integrase mediated transgenesis tool for human genome engineering.....	63
Figure 3.2. An illustration depicting <i>in vitro</i> intra-molecular <i>attH4X</i> x <i>attP4X</i> recombination of a substrate plasmid.....	65
Figure 3.3. SDS-PAGE gel image depicting <i>E. coli</i> cell extract of an induced culture overexpressing C-terminally histidine-tagged Int-h/218 and Int-C3	65
Figure 3.4. Generation of the seamless vector in the presence/absence of scIHF2.	66
Figure 3.5. Representative analysis of an <i>in vitro</i> intra-molecular Int-h/218	67
Figure 3.6. Partial purification of C-terminally histidine-tagged Int-h/218.....	68
Figure 3.7. Recombination activity of partially purified Int-h/218.....	69
Figure 3.8. Int-h/218 mediated intramolecular recombination on a linear substrate plasmid	70
Figure 3.9. <i>In vitro</i> intramolecular recombination by Int-h/218 with <i>attH4X</i> x <i>attP4X</i> vs <i>attL4X</i> x <i>attR4X</i>	71
Figure 3.10. Experimental strategy for <i>in vivo</i> generation of a seamless vector in <i>E. coli</i> strains. ...	72
Figure 3.11. Generation of the seamless vector <i>in vivo</i> from <i>E. coli</i> strains LZ41 and LZ54	75
Figure 3.12. Optimization of culture condition in <i>E. coli</i> strains LZ41 for <i>in vivo</i> seamless vector production	76
Figure 3.13. Optimization of culture condition in <i>E. coli</i> strains LZ41 for <i>in vivo</i> seamless vector production	77
Figure 3.14. Generation of the seamless vector <i>in vivo</i> from <i>E. coli</i> strains LZ41 on large scale.....	78
Figure 3.15. Targeting of seamless vector to <i>LINE-1</i> elements in HT1080 cells	79
Figure 3.16. PCR analysis revealing the integration of seamless vector <i>attL4X-PGKssPuro</i> at <i>attH4X</i> site in <i>LINE-1</i> elements.....	80
Figure 3.17. Screening for <i>attH4X/attL4X</i> recombination events in HT1080 clones.....	83
Figure 3.18. Seamless vector genomic insertion site mapping via deep sequencing.....	86
Figure 3.19. A schematic illustration showing <i>in vivo</i> intramolecular recombination between <i>attH4X</i> and <i>attP4X</i> in substrate plasmid.....	87

Figure 3.20. Screening for <i>attH4X/attL4X</i> and <i>attH4X/attP4X</i> recombination events in HT1080 cells.....	91
Figure 3.21. Screening for <i>attH4X/attL4X</i> recombination events in hES clones.....	95
Figure 3.22. Structure and function of Chimeric antigen receptors (CAR).....	98
Figure 3.23. Targeting of a two-reporter seamless vector <i>attL4X-PGKssPuro-EF-CAR</i> in hESCs.	100
Figure 3.24. Anti-CD19 CAR expression from seamless vectors in hESCs	102
Figure 3.25. Western blot showing anti-CD19 CAR expression.....	104
Figure 3.26. Southern blot analysis of the targeted hESC clone#19.....	105
Figure 3.27. Long-term and stable anti-CD19 CAR expression from seamless vectors in hESCs..	106
Figure 3.28. Strategy for biologics production platform in HEK293T cells	109
Figure 3.29. Screening for seamless vector integration at <i>attH4X</i> site of landing pads in HEK293T cells	113
Figure 3.30. EGFP expression of targeted HEK docking clones.....	113
Figure 3.31. Southern blot analysis of parental and targeted HEK docking clones	116
Figure 3.32. Strategy for therapeutic protein FVIII production platform in HEK293T..	118
Figure 3.33. PCR analysis for <i>pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X</i> integration at <i>attH4X</i> site of landing pads in HEK293T cells.	119
Figure 3.34. Zsgreen and factor VIII expression from HEK bulk culture targeted with FLF8 vector	121
Figure 3.35. Domain structure of CTCF: the three major domains	125
Figure 3.36. Hierarchy of chromosome structures.....	126
Figure 3.37. Cohesin structure and DNA loop extrusion model.....	126
Figure 3.38. CTCF binding sites.	127
Figure 3.39. UTF1 reporter construct (UE)	127
Figure 3.40. Schematic representation of seamless vectors containing CTCF sequences.....	128
Figure 3.41. Schematic representation of seamless CTCF vectors having <i>EGFP</i> cassette replaced with <i>mNeon</i>	128
Figure 3.42. mNeon expression of hES clones targeted with seamless CTCF vectors and negative control vector.	131
Figure 3.43. Strategy for seamless vector transgenesis in human <i>LINE-1</i> elements	132
Figure 3.44. Inverse PCR analysis for identification of the genomic location of seamless vector transgenesis	133
Figure 3.45. Structure of <i>LINE-1</i> element.....	136
Figure 3.46. Retrotransposition cycle of <i>LINE-1</i> element.	137

Figure 3.47. Strategy for seamless vector transgenesis in actively transcribed human <i>LINE-1</i> elements	138
Figure 3.48. Inverse PCR analysis for identification of the genomic location of seamless vector transgenesis	139
Figure 3.49. Seamless vector genomic insertion site mapping via sequencing	140

LIST OF TABLES

Table 1. Comparison of properties of viral vectors, plasmids and transposons (38-40).....	16
Table 2. Comparison of ZFNs, TALENs and CRISPR/Cas9 tools	20
Table 3. List of Primers (5'-3').....	52
Table 4. PCR parameter for GoTaq Flexi DNA polymerase and Q5 High-Fidelity DNA Polymerase.	56
Table 5. Summary of transfection in HT1080 cells with <i>attL4X-PGKssPuro</i> and Int expression plasmids.	79
Table 6. List of genomic coordinates and locations of targeted <i>attH4X</i> in HT1080 bulk cultures ...	81
Table 7. List of genomic coordinates and locations of targeted <i>attH4X</i> in HT1080 clones.....	84
Table 8. List of genomic coordinates and locations obtained upon targeting of either substrate plasmid or seamless vector in HT1080.....	92
Table 9. Summary of transfection in hESCs with <i>attL4X-PGKssPuro</i> and Int expression plasmids.	93
Table 10. Location of targeted <i>attH4X</i> sequences in hESC clones	95
Table 11. Location of <i>attH4X</i> sequences in hESC clones targeted with anti-CD19-CAR seamless vector.....	103
Table 12. Summary of Southern blot analysis for targeted clone #19 using probes complementary to <i>CAR</i> and <i>Puro</i>	105
Table 13. List of 27 “hotspot” <i>LINE-1</i>	107
Table 14. Location of <i>attH4X</i> sites of the landing pads of targeted HEK docking clones.....	113
Table 15. Estimated yield of factor VIII protein from HEK targeted bulk culture and HEK targeted sorted culture based on the factor VIII protein standard curve.....	122
Table 16. Summary of targeting events for FLF8 vector transgenesis in HEK docking clone #3.....	123
Table 17. Summary of fold change for secreted FVIII protein activity from clones obtained from FLF8 vector transgenesis in HEK docking clone #3.	123
Table 18. Summary of the number of mNeon ⁺ clones and their corresponding mNeon ⁺ cells (%) from hESCs targeted with seamless CTCF vectors and negative control vector.	129
Table 19. List of genomic coordinates of hES clones targeted with seamless CTCF vector identified using inverse PCR.	134
Table 20. Summary of targeting events for CTCF vector and negative control vector transgenesis in hESCs and their corresponding cell percentage for mNeon expression.	135

ABSTRACT

Genome engineering is an important component of gene/cell therapy and molecular medicine. In this respect, several DNA editing tools, such as ZFNs, TALEN and CRISPR-Cas system are being currently used for transgenesis. However, issues like insertional mutagenesis, off-target events, and small payload size have affected their usage as precise genome editing tools. A novel λ -Integrase based tool was recently developed in the laboratory of Professor Peter Dröge, where a λ -integrase variant could catalyze conservative site-specific integration of the large payload into a safe harbor site(s) of the human genome. The identified safe harbor site(s), termed *attH4X*, is present within human Long INterspersed Elements-1 (*LINE-1*) in about 1000 copies dispersed throughout the human genome (1). One of the major requirements of this tool is a circular target vector for Int-mediated recombination into genomic *attH4X* for which the traditional *Escherichia coli* is used to produce a plasmid that carries *attP4X*. This requires the inclusion of bacterial backbone along with payload in the target vector construct that also get inserted into the human genome. However, the presence of bacterial sequences in the conventional plasmid reduces its overall efficiency by increasing its size, silencing transgene expression and activating immune responses in the host cell.

Thus, to overcome these limitations and further to improve the safety and targeting efficiency of this tool, the λ -Integrase platform was utilized to eliminate such bacterial sequences by seamless vector generation via *in vitro* intramolecular recombination. For this, a λ -integrase variant protein was partially purified and a protocol for *in vitro* seamless vector generation was established. Additionally, to refine and scale up the production of seamless vector, an *in vivo* platform was developed using *E. coli* strain LZ41 that can produce a high amount of the seamless vector. λ -Int mediated seamless vector transgenesis into human *LINE-1* element has been successfully validated in HT1080 cells and human embryonic stem cells. The switch from plasmid to seamless vector for λ -Int mediated site-specific genomic integration drastically increased the targeting efficiency and reduced off-target events. Furthermore, the seamless vector was utilized for successful transgenesis and expression of therapeutic protein like anti-CD19 chimeric antigen receptor (CAR) in human embryonic stem cells (hESCs). The targeted hES clone showed stable long-term expression of the therapeutic anti-CD19 CAR protein, demonstrating the potential utility of λ -integrase mediated seamless vector transgenesis platform in therapeutic application. Moreover, certain *LINE-1* loci were identified as preferred insertion sites for Int-mediated seamless vector transgenesis. DNA sequence logo on these loci showed specificity in the target site sequence for Int-C3 mediated *attL4X* recombination. It can be further utilized to engineer Int variants for better recognition of this subset of *attH4X* sequence.

To broaden the utility of the platform, the tool was further explored in the field of biologics testing and production. To achieve this, a master cell line of HEK293T was generated by random insertion

of landing pads carrying 4X *attH4X* sequence with a reporter cassette. The landing pad of the master cell line was targeted with a seamless vector encoding a second reporter cassette. This resulted in the high reporter cassette expressing clones that harboured seamless vector in one of the four *attH4X* sequences of the landing pad. The master cell line was further utilized to integrate and express a large transgene of size ~14kb that encodes therapeutic factor VIII protein. The targeted bulk HEK293T showed a high expression of factor VIII protein with a yield of ~2.2µg/ml from 10000 cells indicating the potential application of this platform in the biopharmaceutical sector. Since this platform could allow integration of transgene(s) at a specific genomic locus, it can be further used to test different variants of a recombinant protein and select the best performing variant. In summary, a λ -integrase mediated seamless vector transgenesis platform was developed that manifests dual capability to generate seamless vectors as well as target the human genome for insertion and expression of large transgene cassette. This platform can be broadly utilized for various molecular medicine, gene therapy and synthetic biology applications.

1. INTRODUCTION

The human genome is composed of billions of DNA bases. The ability to precisely alter the genome, also known as “genome editing”, at a specific location has incredible values in molecular biology and biotechnology. Genome editing refers to the process of targeted modification of the genome of organisms that involves DNA insertion, deletion, repair and replacement at a specific site in the genome using molecular scissors. Genome editing has wide applications ranging from therapeutic purposes such as gene therapy, disease treatment, synthetic biology, functional genomics and biologics production. Genome editing can be operated in three different forms that involve, knockout, knock-in and transgenesis of exogenous DNA at safe harbour site of the genome (2,3). Knockout refers to a genetic technique where a specific gene is removed out of the genome to study the gene function by investigating the effect of the gene loss. While in knock-in, a specific gene is replaced with another gene instead of deleting it. Hence, it is considered as targeted insertion where a gene is inserted at a specific locus that allows restoration of the function of the defective gene and also can be used to create disease models (2,3). Besides gene knock-in and knockout, the third category of genome editing involves transgenesis of exogenous DNA at safe harbour site of the genome (4). The transgenesis adds a new gene to the genome, thus it is used for expression of therapeutic proteins, DNA-based vaccines, gene therapy and generation of transgenic plants and animals. (5). The current tools available for genome editing can be broadly classified into two groups based on their mode of action: (a) random insertion of transgene by plasmids, viral vectors, and transposons, (b) site-specific integration of transgene using engineered nucleases (ZFNs, TALENs, and CRISPR/Cas9) or site-specific recombinases (Cre/Flp recombinases, PhiC31 integrase, BxB1 integrase, λ -Integrase) (6,7).

1.1. Random Integration Tools

Plasmid vectors are the most commonly used tools to deliver a transgene into the genome. It is achieved by chemical agents, such as cationic or polymer lipids or physical agents that includes electroporation and microinjection (8). After DNA uptake, the plasmid remains as an episomal DNA in the cytoplasm up to several days. Only a fraction of it enters the nucleus and randomly integrate into the genome (9). Gradual dilution and degradation of the episomally remaining plasmid inside the cell due to continuous cell division, activation of the immune response, insertional mutagenesis, and silencing of transgene expression are major limitations of this tool (8,10,11).

To improve the rate of transgene insertion into the genome, transposons have been developed as a genome-editing tool. The transposon system consists of transgene flanked by the inverted terminal repeats (ITRs) and the transposase-encoding gene, for random integration into a genome (12,13). Examples of commonly used transposons for genome editing include Sleeping Beauty (SB) (14,15) and piggyBac (PB) transposon (12,16,17). However, due to the random transposition of the

transgene, transposon-based genome editing can lead to oncogenesis, insertional mutagenesis and dysfunctionality of the endogenous genes (18,19). Besides, several studies had shown that the transposition efficiency decreases with an increase in the transgene of size beyond 8kb (20). However, recently the payload size of *Sleeping Beauty (SB)* transposon has been increased to ~10kb with a moderate transposition efficiency (20) and *piggyBac (PB)* transposon has been modified to carry a payload size up to ~100kb for *in vitro* murine and human models studies (21-23).

For effective delivery of transgenes into a genome, another commonly used tool is the recombinant viral system. In this system, a recombinant viral vector is constructed by substituting a set of the viral gene responsible for viral replication and pathogenicity with the transgene cassette. The transgene cassette is flanked by terminal sequences and the packaging signal. In addition, a viral packaging construct is also prepared that contains genes encoding structural and replicational proteins. Both these constructs are introduced in a packaging cell line where packaging construct expresses proteins for replication of the transgene cassette and its encapsulation into virus particles to generate the recombinant viral vector (24). Gene transfer using viral vectors results in a high and long-term expression of the transgene (25). Examples of efficient recombinant viral vectors are adenoviral, adeno-associated viral (AAV), herpes simplex viral (HSV), retroviral and lentiviral vectors (26-30). While adenoviral are non-integrating vectors that remain extrachromosomally inside the cells; AAV vectors can persist as episomes within cells and can also integrate in a site-specific manner at AAVS1 on chromosome 19 (31). The packing capacity of adenovirus is 7.5kb of exogenous DNA, while the AAV allows packaging of 4kb of foreign DNA only (32-34). In contrast, HSV can give a leeway of 30kb of foreign DNA for packaging and allows long-term transgene expression (35). Besides all these vectors, retrovirus represents the gold standard vector for gene therapy applications. It can accommodate up to 8kb of foreign DNA (36). However, it lacks the capability to infect non-dividing cells in contrast to lentiviral vectors which can infect both dividing and non-dividing cells. Lentivirus possesses the same packing capacity as retrovirus and can integrate transgene into the host genome for long-term expression (37). The limitations, such as genotoxicity, immunogenicity against the viral component, insertional mutagenesis, and smaller transgene insertion have limited the applications of viral vectors as gene transfer tools in the human genome (25). The comparison among different vectors for random integration is shown in table 1.

Table 1. Comparison of properties of viral vectors, plasmids and transposons (38-40)

Properties	Adenovirus	Adeno-associated virus (AAV)	Retrovirus	Lentivirus	Plasmid	Transposon
Payload capacity	7.5kb	4kb	9kb	8kb	unlimited	10kb
Genomic integration	episomal	episomal/integration	integration	integration	integration/episomal	integration
Efficiency	moderate	variable	low	Low	low	moderate
Transgene expression	transient	stable/transient	stable	Stable	transient	stable
Cell type	broad	broad	dividing cells	Broad	broad	broad
Immunogenicity	high	low	low	Low	low	low
Advantages	high transduction efficiency, higher titre viral stocks, infect most of the cells	less pathogenic and immunogenic, long-term expression, infect most of the cells, high viral titer	high transduction efficiency, long-term expression	suitable for most of the cells, long-term expression	less immunogenic, safe, cheap	high expression, high copy number
Limitations	toxic, inflammatory and immunologic response, low packaging capacity, short-term expression	insertional mutagenesis, small packaging size, requires a helper virus, labour-intensive	insertional mutagenesis, requires dividing cells, low titre virus, activate oncogenesis, less efficient	insertional mutagenesis, activate oncogenesis, less efficient	less efficient in gene transfer, short-term expression, transgene silencing	insertional mutagenesis, low payload size.

With the random insertion of the transgene in the genome, the expression of the transgene is susceptible to silencing due to repressive epigenetic marks of the neighbouring genomic regions (41). Hence, transgene insertion at a transcriptionally active safe harbor site is important to maintain the functionality of the transgene and host genome (42). Thus, site-specific insertion of a transgene could be advantageous over random integration.

1.2. Site-Specific Integration Tool

The site-specific genome editing is commonly achieved by using programmable nucleases such as meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector-based nucleases (TALEN) and clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 system (43). While meganucleases, ZFNs and TALENs achieve DNA editing via direct protein-DNA interactions, CRISPR-Cas9 uses a short guide RNA to specifically target DNA via base pairing and protein-DNA interactions. Programmable nucleases achieve precise genome editing by creating DNA double-strand breaks (DSBs) at a specific location in the genome. These breaks activate the endogenous DNA repair machinery that allows site-specific modification in the genome as shown in figure 1.1 (44,45).

DNA double-strand breaks are repaired by direct re-ligation of DNA at the breakpoint by an error-prone mechanism known as Nonhomologous End Joining (NHEJ). It induces insertions,

deletion, inversion and precise base mutations or substitutions that bridges the break site (46) (figure 1.1). Such indels might lead to frameshift mutations in the coding sequence of the gene, thus resulting in gene disruption. NHEJ is the most prevalent and efficient repair pathway that happen at DSBs at all stages of the cell cycle. This repair pathway involves multiple enzymes and several different outcomes are possible with this repair machinery as shown in figure 1.1 (47-49). Although NHEJ is mostly recognized for gene knockouts, several studies had shown that it can be also utilized for gene knock-ins (50,51). Suzuki *et al.* demonstrated an NHEJ based homologous independent targeted integration (HITI) strategy for an efficient DNA knock-in in both dividing and non-dividing cells at *in vitro* and *in vivo* levels. This strategy was successfully used in a rat model of the retinitis pigmentosa to improve its visual function (52).

Figure 1.1. Different outcomes of genome editing by NHEJ and HDR. Following the DSBs, NHEJ can result in deletion, insertion and inversion of the gene, while HDR can lead to gene addition, gene correction and gene replacement [reprinted with permission (43)].

The second type of repair mechanism that happens in the presence of a homologous DNA template is homology-directed repair (HDR). The HDR machinery uses the exogenous DNA template to synthesize new DNA thus incorporating the template at the break site that can lead to indels and base correction as shown in figure 1.1 (53-56). Such a repair is important for genome editing to achieve gene knock-in, gene knockout, gene replacement and point mutations. The sequence of the homologous DNA or sister chromatids is crucial for HDRs. Thus, the HDRs mainly occur at the late S and G2 phases where sister chromatids act as homologous DNA template for the DSBs repair (49,57,58). HDRs are known to be high fidelity processes; however, it is less frequent as repair machinery as compared to NHEJ. Although, the efficiency of the HDR pathway is limited in non-dividing cells, yet it can allow targeted genome modifications that can have a significant impact in the field of gene therapy (57).

1.2.1. Meganucleases

Meganucleases (also known as homing endonucleases), were the first nuclease recognized as DNA genome engineering tool (45,59,60). These are 20-37kDa site-specific nucleases that are active as monomers and can recognize long target sequences (14-40bp). Meganucleases are different from restriction endonucleases by their longer target sequences which rarely occur in the genome. These nucleases first recognize the target sequence as a monomer, followed by endonucleolytic cleavage to generate a double strand break to promote gene targeting via HDR (61). As DNA editing tool, meganucleases are less toxic as they are naturally occurring, and it can generate site-specific cleavage. However, their applications in functional genomics and therapeutic aspects are limited due to low frequency of target sequence in the genome and their tendency to tolerate degeneracy in the target sequence leading to off-target cleavage (62). To widen their application in genome engineering, the naturally occurring forms of meganucleases have been engineered to recognize diverse target sites in the human genome (62). These have been fused to the protein domain of I-DmoI and I-CreI to create chimeric meganucleases to improve site-specific cleavage of target sequence (63). However, their utility is highly dependent on the efficient strategies to engineer meganucleases to target the desired sequence which acts as a major challenge for their applications. Recently, a fusion of meganucleases with TALEN, known as MegaTALs, was developed that can perform a high level of targeted DNA modifications in human cells (64). Besides, meganucleases have also been fused with DNA end-processing enzymes to achieve error-prone NHEJ to increase the chances of mutagenic events at the target site (65).

1.2.2. Zinc finger nucleases (ZFNs)

Zinc finger nucleases (ZFNs) are the first generation engineered nucleases that consist of DNA binding domain Cys2-His2 zinc finger protein and FokI restriction endonucleases as cleavage domain (66). It functions as dimer where a zinc-finger DNA binding domain binds to half of the specific site on DNA (9-18bp). This is followed by FokI restriction endonucleases cleavage in the spacer sequence (5-7bp) which separates two flanking zinc-finger binding sites as shown in figure 1.2 (67). ZFNs usually consist of 3-4 zinc-finger DNA binding domains recognizing DNA triplet where each domain is made up of ~30 amino acid residues and stabilized by Zn^{2+} ion (68). ZFNs have been successfully used to genetically modify human embryonic stem cells (hESCs), induced pluripotent stem cells (iPSCs), mouse and zebrafish models (69-72). It has been applied for therapeutic approaches for gene correction of Hemophilia B and SCID in mammalian cell lines and mouse models (73,74). However, ZFN mediated genome editing is associated with issues like off-target cleavage (75,76), targeting of the non-G-rich sequences (77) and also the requirement to construct custom-designed zinc-finger arrays each time (78) limit its downstream applications.

Figure 1.2. Schematic diagram of the ZFN system consisting of Zinc finger protein and FokI endonuclease [reprinted with permission (79)].

1.2.3. Transcription activator-like effector-based nucleases (TALEN)

Transcription activator-like effector-based nucleases (TALENs) are second generation nucleases because of their modular structure and function. It comprises of an amino-terminal TALE as DNA binding domain (~34 amino acid residues) that recognizes DNA sequence and FokI as cleavage domain (figure 1.3) (80,81). Like ZFNs, TALEN also functions as dimer where a TALEN monomer recognizes a specific sequence of 14-20bp followed by FokI cleavage within a spacer sequence (12-19bp) separating two flanking TALE binding sites (81). As compared to ZFNs, TALEN array is easy to construct, shows higher genome editing specificity and reduced toxicity (82). TALEN has been used to correct the genetic error for sickle cell disease (83,84), epidermolysis bullosa (85) and xeroderma pigmentosum (86). Besides gene correction, TALEN has been used to generate gene knockout in *C. elegans* (87), mice (88) and zebrafish to develop disease models (71). However, the large size of each TALEN limits *in vivo* delivery; also, its repetitive DNA sequences cause unwanted recombination events (77).

Figure 1.3. Schematic diagram of the TALEN system consisting of engineered TALE protein and FokI endonuclease [reprinted with permission (79)].

1.2.4. Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system

CRISPR/Cas9 system is a third-generation genome editing tool that has been utilized in a wide variety of applications. This tool has been developed from the clustered regularly interspaced short palindromic repeats (CRISPR) Type II bacterial immune system (89-91). The tool comprises of two components: a) short synthetic guide RNA (gRNA) that carries ~20 nucleotide targeting

sequences followed by a downstream protospacer-associated motif (PAM) motif (2-5 nucleotides) (b) CRISPR-associated endonuclease (Cas9) which recognizes and cleaves the DNA near to the PAM motif (figure 1.4). The sgRNA directs the CAS9 endonuclease to the target sequence in the genome to create a double-strand break which is repaired by either NHEJ or HDR. This tool can be engineered to alter the genomic target by simply modifying the target sequence on the gRNA (92-94). As compared to ZFNs and TALEN, this tool is much simpler, can be easily manipulated and provide high efficiency in generating targeted mutations (95). The comparison between all these three tools is shown in table 2.

Table 2. Comparison of ZFNs, TALENs and CRISPR/Cas9 tools [information taken from (96-98)]

Advantages	smaller size of ZFNs allows efficient packing in viral vectors.	high specificity, does not require linkages between TALEN motifs	simultaneous targeting of multiple genes
Limitations	requires cloning to include linker sequence for fusing zinc finger motifs, target site size is limited to multiple of three, results in off-target events	difficult to pack in virus system due to large size, undesirable recombination events due to repetitive sequence of the TALEN array	difficult to pack in the viral system due to large size of Cas9 (4.2kb), limitation of PAM sequences in the human genome, generation of indels in the genome

There are different variants of CRISPR/Cas9 nucleases that had been developed as genome-editing tools as shown in figure 1.5. The first one includes the wild-type Cas9 that can site-specifically create a double-strand break to induce NHEJ or HDR machinery for indels generation at the breakpoint as shown in figure 1.5 a (99). The second type of Cas9 variant includes a mutant form, called as Cas9^{D10A} that has only nickase activity. The Cas9^{D10A} form can cleave only the target strand of DNA, thus, there is no activation of NHEJ (figure 1.5 b). However, in the presence of homologous DNA, the DNA is precisely repaired by the high-fidelity HDR pathway thus reduces the generation of indels mutations (100). This tool is best suited for site-specific targeting. It can be achieved by applying a paired Cas9 complex to generate staggered DNA nicks, followed by DNA repair by the HDR or NHEJ pathway for site-specific insertion (101). The third variant includes a nuclease deficient Cas9 (dCas9) that can just bind to DNA without cleaving it (figure 1.5 c). It has been used in fusion with effector proteins to activate or silence a gene (102). The dCas9 was also fused with EGFP in order to visualize DNA at multiple regions (103).

The fourth variant that has been recently developed includes a prime-editor protein (PE) which is a fusion of Cas9 H840A nickase (a mutant variant that can cleave only one strand of target site) and a pentamutant of M-MLV reverse transcriptase (RT) (figure 1.5 d) (104). The prime-editor protein forms a complex with a prime-editing guide RNA (pegRNA) consisting of the desired edit and along with a homologous sequence of target site that acts as a primer binding site (PBS). The pegRNA directs the complex to the target site where the Cas9 protein introduces a single-strand cleave on the target site. It is followed by hybridization of the pegRNA with the 3' end of the nicked DNA strand. This act as a template for reverse transcription of the desired edit which is directly polymerized on the nicked target DNA strand (figure 1.5 d). The original DNA strand is replaced by the edited DNA strand that results in the formation of a heteroduplex with one edited and one unedited strand. Finally, the complex is resolved by nicking the non-editing strand with a different guide RNA that activates the cell's endogenous DNA repair machinery to replace the unedited DNA strand by using the edited strand as the template (104). The prime editing system offers advantages over conventional CRISPR–Cas9 system as it only introduces single-strand breaks, uses DNA mismatch repair machinery and reported to show fewer undesirable off-target and on-target effects (104).

A fifth variant was developed that can edit a single nucleotide without creating any double-strand breaks. There are two classes of base editors - cytosine base editors (CBEs) and adenine base editors (ABEs). It includes the fusion of Cas9 nickase or dCas9 with cytidine deaminase (figure 1.5 e). The base editor is guided to target sequence by gRNA where it can convert cytidine to uridine near the PAM site. The uridine is then changed to thymidine through base excision repair, thus, generating a C to T alteration (or a G to A on the reverse strand). Similarly, to generate adenine deaminase, modified *Escherichia coli* TadA domain, was fused to Cas9 protein that can create a change from A to G (or T to C on the opposite strand) (figure 1.5 e) (105-107). While most of the base editor is designed to make base modification in a narrow window near to the PAM sequence, several base editing systems are developed that can function in a wider window creating a wide spectrum of single-nucleotide variants, thus, suitable for directed evolution applications.

CRISPR/Cas12a (also known as Cpf1) is a third-generation genome editing tool which is a new addition to the CRISPR toolbox. It is an RNA-guided tool analogous to CRISPR/Cas9, however, it has unique features that distinctly differ from the CRISPR/Cas9 system (figure 1.5 f). Cas12a is a smaller size protein compared to Cas9 and it recognizes a PAM sequence that is widely available in the human genome (108). Moreover, the PAM sequence for Cas12a is T-rich at the 5' end in contrast to G-rich PAM sequence for Cas9 protein. Different variants of Cas12a have increased its range of target sequence in the genome. Cas12a creates a staggered double-strand break at a PAM distal region which makes it more advantageous for gene knock-in applications (109,110). All these unique features introduced the CRISPR/Cas12a system as a promising genome engineering tool that has been utilized in gene targeting, base editing and transcriptional and epigenetic modulations. It has been used in various organisms starting from prokaryotes to the human. The lower off-target events with CRISPR/Cas12a system has improved the CRISPR toolbox in terms of safety concerns (111,112).

(a) Genome Engineering With Cas9 Nuclease

(b) Genome Engineering By Double Nicking With Paired Cas9 Nickases

(c) Localization With Defective Cas9 Nuclease

(d)

(e)

(f)

Figure 1.5. Different variants of CRISPR/Cas9 tools: (a) wild-type Cas9 creating DSBs for NHEJ or HDR mediated repair (b) paired mutant variant, Cas9D10A, generating single-strand nick to generate a staggered DSBs (c) The nuclease deficient Cas9 (dCas9) fused with effector protein for transcriptional regulation and fluorescence tagging (reprinted from www.neb.sg/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology) (d) Prime editing complex comprised of Cas9 nickase, RT and pegRNA, generates single-strand cleave at PAM-containing strand that results in hybridization of PBS with 3' end of nicked strand. RT uses it as a template for reverse transcription of the desired edit on the nicked DNA [adapted with permission (104)] (e) Molecular reactions underlying cytosine base editing (CBE) and adenosine base editing (113): the CBE catalyzes the conversion of cytosine to uridine and the ABE catalyzes the conversion of adenosine to guanine via Base excision repair (BER) or Mismatch Repair (MMR) [reprinted from www.genengnews.com/insights/all-about-that-base-editing] (f) Schematic representation of CRISPR/Cpf1 system composed of CRISPR RNA (crRNA) and the Cpf1 endonuclease [reprinted with permission (79)].

The CRISPR/Cas9 tool is suitable for multiplex editing by designing multiple sgRNAs with different target sequences that can be used simultaneously in the cell. Due to its simplicity and high efficiency, currently, the CRISPR/Cas9 system is widely used in almost every application of genome engineering (114). It has been widely used in knockout studies, selectively repressing or activating target genes, RNA editing, designing disease model, creating gene drive, biomedicine, and tagging of DNA using fluorescence markers for microscopic visualization as shown in figure 1.6. The versatile nature of the CRISPR/Cas system has scaled up its usage in genome editing technologies to target different genes in several organisms, introducing single point mutations, large deletions or genomic rearrangements, base editing, targeting protein domains for epigenetic modification or transcriptional regulation and manipulating chromatin topology (115-123).

Figure 1.6. Different applications of CRISPR/Cas9 technology beyond genome engineering. It includes manipulation of gene regulation, epigenetic modifications and chromatin topology. Additionally, it can also be used for chromatin imaging, RNA targeting and base editing [reprinted with permission (95)].

Even though the CRISPR/Cas9 system has been utilized in multiple applications, the system was found to be associated with several shortcomings (124-127) :

- a) Insufficient indels and low HDR frequency: the generation of indels using CRISPR/Cas9 varies from different target sites. It has shown unsurpassed efficiency to generate indels in a variety of cells as compared to engineered nucleases. Besides, the efficiency of the HDR pathway for transgene insertion had also been found low with this tool (114).
- b) Off-target issue: CRISPR/Cas9 system has been reported to generate unintended nonspecific DNA modifications, such as large indels, chromosome aberrations, gene rearrangement, through sgRNA dependent/independent manners. Even base editors also showed unpredictable off-targets in a sgRNA independent manner (114).
- c) Large size in gene delivery: due to the large size of Cas9 protein, it is difficult to pack both sgRNA and Cas9 protein in single virus particle, thus limiting *in vivo* delivery of the whole complex (114).
- d) Lack of PAM sequence: each Cas9 protein has its own preferred PAM sequences. Recognition of the PAM sequence is necessary for Cas9 mediated DSB. In some cases, the CRISPR/Cas9 system is limited to make DNA modifications at the target site due to the unavailability of the PAM sequence (114).
- e) Immune response: since the CRISPR/Cas9 system utilizes bacterial protein Cas9 in eukaryotic cells for genome editing, it has been reported to elucidate immune response and cellular

toxicity. Studies have shown that use of this system *in vivo* has resulted in the generation of antibodies against Cas9 protein (114).

1.2.5. CRISPR-associated Transposase (CAST)

A new tool for genome editing was recently discovered that combines CRISPR and transposase system, termed as, CRISPR-associated transposase (CAST) (128). It utilizes catalytically inactive Cas effector protein along with three transposase subunits. Strecker *et al.* reported utilisation of Cas12k-guide RNAs along with of transposase subunits TnsB, TnsC and TniQ for site-specific insertion of genetic cargo (figure 1.7 a) (129). Similarly, Klompe *et al.* utilized multiple complexes composed of Cas 6,7,8 effectors along with tnsA/B/C and TniQ transposase subunits for targeted insertion of a transgene at the specific genomic loci (figure 1.7 b) (130). The site-specific integration by these techniques occurred at 49–66bp downstream of G- and T-rich PAM (129) and 46-55bp downstream of C-rich PAM (130). The integrated DNA was flanked by 5bp target site duplications which represent a hallmark of Tn7 transposon integrations. CAST mediated DNA integration doesn't induce any DSBs in the genome and ensures more safe and specific integration. It can insert transgene of size 500bp with high efficiency (~60%) and shows a moderate insertion rate of ~25% for cargo size of 10kb (129). However, CAST has only been validated in *E. coli* genome and was found to be associated with significant off-target insertions scattered across the genome (129).

Figure 1.7. CRISPR-associated transposase (CAST) developed by (a) Strecker *et al.* (129) (b) Klompe *et al.* (130) [reprinted with permission (131)].

1.2.6. Site-specific integration by recombinases

As I have outlined above, engineered nucleases have been utilized as transgenesis platform where the transgene is being integrated into the target site by the HDR pathway. However, this pathway is often inefficient, and its efficiency varies with different target sites. Hence, site-specific recombinases have been used as an alternative pathway to achieve transgenesis (132,133). The site-specific recombinases

catalase the transgenesis reaction by cleaving, exchanging and rejoining the DNA strand after binding to short target sequences. It results in different outcomes that include integration, excision and inversion of the DNA segment flanked by the target sites. A promising feature of site-specific recombinases is that the recombination carried out by them is highly specific, efficient and fast (134,135). There are two types of site-specific recombinases based on the conserved nucleophilic amino acid residue: serine-based recombinases (e.g. gamma-delta resolvase, ϕ C31 integrase, Tn3 resolvase, Hin recombinases, Bxb1 integrase) that uses a serine residue to cleave DNA, and tyrosine-based recombinases (lambda-integrase, Cre recombinase, FLP recombinase) that uses tyrosine to cleave DNA in the process of recombination (135-137).

The common mechanism of site-specific recombinases for recombination (figure 1.8) involves the following steps:

- (a) recognition of the target sequences in two DNA molecules that are to be recombined
- (b) the protein-bound DNA sites are then brought together (synapsed)
- (c) the DNA strands are cleaved at fixed points
- (d) the DNA strand exchange is accomplished by a transient covalent link between the hydroxyl group of the nucleophilic tyrosine/serine residue of recombinase with the phosphate group of the DNA strand
- (e) exchanged DNA strands are ligated back to form the recombinant products.

While the serine recombinases introduce a pair of double-strand breaks at the two DNA molecules to be recombined, the tyrosine recombinases exchange one pair of single-stranded DNA strands at a time via formation of an intermediate DNA structure (Holliday junction) as shown in figure 1.8 (135,138).

Figure 1.8. Schematic representation of site-specific recombination by serine and tyrosine recombinase (reprinted from <https://slideplayer.com/slide/7463425>).

1.2.6.1. Types of serine recombinases

- i) BxB1 integrase is a serine recombinase expressed by mycobacteriophage Bxb1 that carry out site-specific recombination between two attachment (*att*) sites: *attP* site (48bp) on phage genome and *attB* site (38bp) on *Mycobacterium smegmatis* genome. This results in the formation of prophage flanked by *attL* and *attR* (figure 1.9). The *attP* site is positioned 25bp upstream of Bxb1 *int* gene flanked by imperfect inverted repeats, while the *attB* site is located at 3' end of the *groEL1* gene flanked by perfect inverted repeats (139). The recombination catalyzed by this integrase is highly selective for the cognate attachment sites and is strongly unidirectional. The excision reaction (*attL* x *attR*) is tightly regulated by the recombination directionality factor (RDF) encoded by Bxb1 *gp47* gene (figure 1.9) (140).

Figure 1.9. Bxb1 integrase mediated integration and excision reaction between *attP* site on phage genome and *attB* site on *Mycobacterium smegmatis* [reprinted with permission (141)].

Bxb1 integrase has been found to be functional in mammalian cells where it can integrate plasmid DNA at chromosomal *attP/attB* site with better efficiency than FLP recombinase (142,143). It can also catalyze *in vitro* recombination between the *attP* and *attB* sites without any cofactors, divalent cations and requirement for the supercoiled form of DNA (139). Bxb1 integrase has been successfully applied in the recombinase mediated cassette exchange (RMCE) system where it was used to site-specifically integrate large transgene cassette that expresses monoclonal antibody in CHO cells (144,145). One major drawback of the system is that the integrase cannot integrate plasmid carrying *attP/attB* sites at endogenous pseudo *att* sites inside the human genome (142), thus, limiting its application as a human genome engineering tool for pre-engineered docking sites.

- ii) The bacteriophage phiC31 integrase is a potential enzyme for site-specific recombination. The phiC31 integrase catalyses recombination between two different target sites, *attP* and *attB*. This results in the formation of two new sites, *attL* in the left junction and *attR* in the right junction flanking the recombinant product. The integrase cannot recognise *attL* and *attR* sites, hence the recombination is unidirectional. The integrase can catalyse the recombination without any auxiliary host factor and in the absence of negatively supercoiled DNA. Both the *att* sites carry an identical 2bp core sequence flanked by inverted, imperfect repeats (figure 1.10). The *attP* sequence is slightly longer than the *attB* sequence (146,147).

Figure 1.10. Nucleotide sequences of the *attP* and *attB* sites recognised by phiC31 integrase. Both the sites contain integrase binding sites (P and P' in *attP*, B and B' in *attB*) and a di-nucleotide core sequence [reprinted with permission (148)].

The phiC31 integrase has been utilized as a human genome integration system where pre-inserted *attP* site in the genome was successfully targeted for site-specific integration of *attB* carrying plasmid. Moreover, a pseudo *attP* site has been identified in the genome (figure 1.11) which is 40% similar to wild-type-*attP* sequence (149). The pseudo *attP* site is located at the highly transcribed region and it could be targeted with 10-fold higher rate by phiC31 integrase for site-specific integration than random sites (150,151). However, the integration event catalyzed by bacteriophage phiC31 integrase results in chromosomal aberrations due to its

recombinase activity at multiple pseudo-*att* sites. It also leads to indels flanking the inserted transgenic DNA, debatable true site-specific recombination events and shows low targeting efficiency (150,152). Other serine recombinases have also been explored as genome editing tools, yet factors like the insertion of the transgene into endogenous safe harbor sites have always been an issue associated with these tools (149). Hence, tyrosine recombinases like Cre-*loxP*, Flp-*FRT*, and λ -Integrase have been explored for site-directed transgenesis.

Figure 1.11. Schematic representation of site-specific recombination by phiC31 integrase between pseudo *attP* site in the genome and *attB* site on donor plasmid (reprinted from www.systembio.com/phic31).

1.2.6.2. Types of tyrosine recombinases

The Cre recombinase (“Cyclization recombinase”) and Flp (“flippase”) recombinase, were derived from phage P1 and yeast 2- μ m circle plasmid, respectively. These recombinases can catalyse reversible site-specific recombination between two identical sequences, 34bp *loxP* sites for Cre recombinase and 48bp *FRT* sites for Flp recombinases (figure 1.12). Both recombinases do not require any accessory protein, specific DNA structure and arm-type binding sites. The physiological functions of Cre/Flp recombinases mainly involve excision and inversion of DNA as compared to typical tyrosine integrases coded by temperate phages (153,154).

Figure 1.12. Nucleotide sequences of 34bp *loxP* site and 48bp *FRT* site recognized by Cre recombinase and Flp recombinase respectively. The sites composed of 13bp inverted repeats for recombinase binding sites and an 8bp spacer region for strand exchange [reprinted with permission (148)].

Due to their simplicity, Cre/*loxP* and Flp/*FRT* systems have been widely used for *in vivo* genome engineering to generate insertions, deletions, inversions, and translocations in the genome (figure 1.13) (155-157). One of the best examples of Cre-lox recombination system application is the “Brainbow” that enable different color labelling of each neuron of mice to create a detailed map of the neuronal network allowing neuroscientists to study neurons and glial cells as a function of time (158). The Cre/*loxP* and Flp/*FRT* systems are widely used in recombination-mediated cassette exchange (RMCE) strategy to engineer the genome of mammalian cell lines for the production of monoclonal antibodies (159,160). However, transgenesis of exogenous DNA is inefficiently catalyzed by these recombinases because they can predominantly catalyze the excision reaction in an *in vivo* condition leading to excision of integrated DNA from the genome. Hence, the system other than RMCE, is not suitable for stable integrative recombination events (153,154). In addition, overexpression of Cre/Flp recombinases for integrative recombination has resulted in cell growth inhibition and undesirable non-specific recombination events in the genome (161,162).

Figure 1.13. Cre recombinase mediated recombination between *loxP* sites results in (a) excision of DNA (b) inversion of DNA (c) translocation of DNA [reprinted with permission (163)].

1.2.6.3. λ -Int-mediated site-specific recombination

The site-specific recombination mediated by λ -Int represents a naturally evolved transgene addition system, which can insert large circular transgene cassette (up to 48kb in the case of the λ genome itself) into a specific genomic target sequence with high efficiency. The integration/excision system mediated by wild-type phage λ for the lysogenic and lytic phage life cycles is shown in figure 1.14. The lysogenic phage life cycle involves the integration of phage DNA into the bacterial genome by using a pair of attachment (*att*) sites: *attB* (in the bacterial genome) and *attP* (in the phage DNA) which is catalyzed by bacteriophage λ encoded integrase protein (164). The integration results in prophage which is flanked by *attL* site on the left and *attR* site on the right. The prophage is fully

compatible for excision recombination to excise the phage genome from the bacterial genome under unfavourable physiological conditions to enter the lytic cycle (164).

Figure 1.14. λ -Int mediated site-specific recombination between *attP* on bacteriophage genome and *attB* on *E. coli* genome leading to the integration of prophage flanked by *attL* at the left junction and *attR* at the right junction [reprinted with permission (165)].

The 241bp *attP* and 21bp *attB* consist of a pair of 7bp inverted core-binding sites which are separated by a 7bp overlap (also called spacer) region. These sites are involved in strand exchange during recombination. The *attP* site has arm-binding sites for Int binding at about 80bp towards the left and 150bp towards the right from the overlap. The Int enzyme comprises an N-terminal domain (1-64 aa residues), a core binding domain (65-169 aa residues) and a C-terminal catalytic domain (170-356 aa residues). The N-terminal domain is responsible for recognizing several arm binding sites, and its DNA binding increases the affinity of the C-terminal domain to the core binding sites. This allows Int to create specific synaptic complexes between *att* sites which catalyze strand exchange via Holliday junction intermediates (166). The binding of wild-type Int at the *attP* arm sites to form functional recombinogenic complexes requires accessory protein co-factors like FIS (factor for inversion stimulation), Xis (excisionase) and IHF (integration host factor) (167).

The heterodimer IHF protein is encoded by *E. coli* genes *himA* and *himD*. It binds to *attP* arm sites to induce DNA bends, which allows the same Int monomers to catalyze both integrative and excisive recombination by simultaneously binding to arm and core binding sites (167). The phage-encoded Xis protein promotes excision by inducing a sharp bend in DNA and thus facilitates *attL/attR* recombination by Int with assistance from IHF (167). FIS is encoded by *E. coli* gene *fis* and it bends DNA to stimulate the excision reaction 20-fold under limiting conditions of Xis (168).

Site-specific recombination mediated by wild-type λ -Int involves strand cleavage, exchange via Holliday junction and re-ligation of the recombining DNA partners as shown in figure 1.15. The first step of integration includes the formation of an intasome between negatively supercoiled *attP*, Int and IHF (169). This complex catches protein-free *attB*, presumably through random collision to

form a synaptic complex (170). Secondly, the top strand at each *att* site is cleaved by Int and forms a covalent phospho-tyrosine linkage by the tyrosine residue at the 3' end of DNA. The free OH group of the partner DNA strand subsequently attacks this linkage that results in a Holliday junction intermediate structure (171). The Holliday structure is then resolved by cleavage, exchange, and ligation of the bottom strands (172). The switch of strand cleavage from top to bottom strand involves isomerization of the Holliday junction that is followed by something akin to proofreading at the identical overlap sequence region in both strands. The completed recombination results in *attL* and *attR* sites that can serve as substrates for excision reaction by Int, XIS, and IHF proteins through the same process as mentioned above (172). Thus, Int-mediated recombination performs an ordered, sequential pair of strand exchange along with the intermediate step of proofreading that helps to limit off-targeting events of recombination.

Figure 1.15. Site-specific recombination mediated by λ -Int via a Holliday junction model involves strand cleavage, exchange and re-ligation of the recombining DNA partners [reprinted with permission (173)].

The bacteriophage λ -Int is one of the best-described recombinases (174,175), but the requirement of negative supercoiling of *attP* and accessory proteins has limited its use in eukaryotic systems (176). In this respect, different mutants of λ -Int were generated mostly by the laboratory of Peter Dröge that can promote recombination in the absence of accessory proteins. The mutants include Int-h (177), Int-h/218 (178), Int-1, Int-5 (179), Int-C2 and Int-C3 (180). The mutants are found to have a higher affinity towards the wild type and non-cognate *att* sites, as compared to the wild-type λ -Int (179,180).

Chandra *et al.* (2016) recently showed the integration system of phage λ as a novel transgenesis tool for human genome engineering (1). This simple transgenesis tool involves highly active, co-factor independent λ integrase (Int-C3 or Int-h/218) that can insert functional transgene cassette into the human genome by site-specific recombination. The identified endogenous human target sequence, termed *attH4X*, is present within human *Long INterspersed Elements-1 (LINE-1)* in about 1000 copies dispersed throughout the human genome (1). It involves intermolecular recombination between two sites, *attH4X* (18bp) found in a small subset of human *LINE-1* and a modified *attP* site, *attP4X* (241bp), placed on circular target vectors. Mutant Int-C3 catalyzes intracellular integrative recombination between genomic *attH4X* and *attP4X* (on a plasmid) resulting in the integration of the entire bacterial plasmid flanked by hybrid *attL4X* and *attR4X* sites into *LINE-1* (figure 1.16). This tool has been exemplified with pluripotency reporter cassettes and constitutively expressed payloads that remain functional in *LINE-1* targeted hESCs and their differentiated progenies (1).

Figure 1.16. A schematic diagram showing λ -Int-mediated recombination between genomic *attH4X* (located in *LINE-1* elements) and *attP4X* (present on a plasmid target vector). The recombination leads to the site-specific integration of the entire plasmid into the human *LINE-1* elements, flanked by *attL4X* (left junction) and *attR4X* (right junction) [reprinted with permission (181)].

One of the requirements for the λ -Integrase transgenesis tool is a circular target vector for Int-mediated recombination into *attH4X*, for which the traditional *Escherichia coli* is used to produce a plasmid that carries *attP4X*. This requires the inclusion of bacterial backbone along with payload in the target vector construct that also get inserted into the human genome (1). The bacterial backbone consists of antibiotic-resistant genes, an origin of replication, potential cryptic expression signals, and unmethylated CpG motifs (182). All these parts of the bacterial backbone are required for plasmid

DNA amplification and selection in bacteria cells; however, they are unnecessary with respect to gene transfer and can be associated with safety issues in clinical applications.

1.3. Seamless vector and its significance

In the context of random transgene insertions using bacterial plasmids, studies had shown that a bacterial backbone in a plasmid can be responsible for the silencing of *in vivo* transgene expression via conversion of bacterial sequences into heterochromatic structures (183). Suzuki *et al.* proposed that the cell has an unknown host defense mechanism that can recognize the bacterial sequences leading to heterochromatin formation (184,185). Furthermore, the bacterial backbone carries unmethylated CpG motifs that can bind to the Toll-Like-Receptor 9 (TLR9) of antigen-presenting cells (APCs). It results in activation of the innate immune system. The binding can result in activation of T-helper type 1 (TH1) cells dominated response by initiating maturation, differentiation, and proliferation of natural killer cells (NK), T-cells and macrophages (186,187). This type of immune response has been prevalent for plasmid DNA vectors used for DNA vaccine and gene therapy. In addition, the size of conventional plasmid DNA is increased due to the presence of about 2kb bacterial sequences. The size of the plasmid vector influences the easy diffusion of the plasmid into the tissue by crossing the cell membrane, escaping lysosome diffusion in the endocytic pathway, entering cytoplasm and then ultimately crossing the nuclear membrane (188-190).

To summarize, the bacterial sequence of the conventional plasmid contributes to a significant portion of the plasmid that declines its overall efficiency by increasing its size, silencing transgene expression, rising biosafety concerns and activating immune responses in the host cell. Hence, removal of bacterial sequences from a conventional plasmid can significantly improve the efficiency and mitigate some of the safety concerns associated with plasmid DNA to be used for non-viral gene transfer (191).

To facilitate and improve the safety of plasmid DNA, it is desirable to remove bacterial sequences from the conventional plasmid. The DNA molecule that is generated by removing the bacterial backbone from a conventional plasmid is known as seamless vector (aka minicircle). Seamless vectors are covalently closed circular (CCC) supercoiled DNA molecules and smaller in size than a conventional plasmid (192,193). Seamless vectors can be generated from a parental plasmid which consists of bacterial backbone along with the gene of interest flanked by two recombination target sequences. It can be produced by *in vivo* and *in vitro* recombination by flanking the gene of interest (GOI) in the parental plasmid with two recognition sites for site-specific recombinases. Upon expression of the recombinase, intramolecular recombination will lead to the formation of two supercoiled circular DNA molecules that are often linked in a catenane: a) the

miniplasmid (MP) carrying undesired bacterial backbone elements and, b), the seamless vector (MC) carrying the GOI (figure 1.17) (191).

Figure 1.17. Intramolecular recombination in the parental plasmid (PP) resulting in two supercoiled circular molecules; minicircle (MC) and miniplasmid (MP). MC and MP can be catenated (reprinted from <https://plasmidfactory.com/products/custom-manufacturing/minicircle-dna>)

1.3.1. Generation of the seamless vector by recombinases

Several site-specific recombinases, such as Cre/FLP recombinase, lambda integrase, PhiC31 integrase and ParA resolvase had been used to generate MCs *in vivo* (192,194-196). Recombination carried out by Cre/FLP recombinases is reversible and bi-directional. This leads to the formation of unwanted dimeric and trimeric MCs along with monomeric MCs. Additionally, the formation of different concatemers of minicircle, miniplasmid and parental plasmid was also observed by Cre/FLP recombination events. Even though mutation in one recognition site had improved the generation of monomeric minicircle, yet the problem of formation of a high amount of concatemers persisted. In contrast, the PhiC31 integrase recombination is strictly irreversible and unidirectional. Minicircle generated using this integrase was found to be highly efficient without the formation of any concatemers (195,197,198). Similarly, the ParA resolvase of the multimer resolution system of the broad host range plasmids RK2 and RP4, can carry out intramolecular recombination only in one direction, thus there was no formation of unwanted multimeric minicircle or concatemers during *in vivo* production (196). Studies had shown that the transformation of low-copy number ParA resolvase plasmid with high copy parental plasmid in *E. coli* achieved only 50% efficiency for minicircle generation. However, upon incorporating the ParA recombinase expression cassette in the parental plasmid, the minicircle generation efficiency was drastically improved to 99.57% (146,196,199-201).

1.4. Overview of biologics production platform from the mammalian expression system

The production of recombinant mammalian proteins is in high demand for both applied research and in the biopharmaceutical sector. While the applied research requires recombinant mammalian proteins for functional and structural studies, the pharmaceutical sector is heavily investing in protein therapeutics i.e. biologics, as a new approach to treat different human diseases (202). Several multi-billion-dollar biopharmaceutical companies are focused on the production and application of recombinant mammalian proteins for therapeutic purposes (203). Most of the research for the

development of expression systems for biologics production is conducted on prokaryotes (204), yeast systems (205,206) and mammalian cell lines. Out of all, mammalian cell lines are mostly preferred for production of therapeutic protein as they can perform posttranslational modifications, protein folding and assembly similar to native human proteins (113,207,208).

The most commonly used cell lines for biologics production are Chinese Hamster Ovary cells (CHO) and Human Embryonic Kidney cells (HEK293-T). For large scale production, these cell lines provide several advantages, such as: a) high transfection efficiency b) robust growth in commercially available serum-free media for the culture c) suitable for both random and targeted integration of transgene d) produce recombinant proteins having human-like post-translation modifications e) easily generate stable cell clones for high yield of recombinant protein expression (128,209). Other cell lines such as Baby Hamster Kidney (BHK), Mouse myeloma NS0 and SP2/0 derived cells and HeLa (Human cervical cancer) cells are also used for mammalian protein production. However, production of protein from mammalian expression systems at a large scale is a laborious, complicated and cost-intensive process (210,211). Even though various platforms have been exploited for biologics production, there is a continuous demand for new robust mammalian expression platforms (212).

In most cases, the transgene expression construct for biologics production is introduced randomly in the mammalian cell genome. This is usually carried out by transposons or viral systems (209). Few examples of naturally-occurring transposons which are used to randomly insert transgene in mammalian cells are Sleeping Beauty (SB) (85), Piggy Bac (PB), the medaka fish Tol2 element and the Tc1/mariner element Mos1 (16). Several cell lines such as CHO, HEK293, CHO lec 1 and HEK-GnTI were successfully modified to overexpress membrane and secreted proteins using PiggyBac (213-215). Besides transposons, lentiviral vectors are also used to modify CHO to produce recombinant proteins (216,217). These vectors can integrate transgene at actively transcribed genomic loci (218) and they show high transducing efficiency in a broad range of cell type. Hence, lentiviral vectors have been successfully applied to BHK and HEK293 cell lines to generate highly productive strains (219,220). However, several limitations have restricted their use in the biopharmaceutical sector. The vector generation is time-consuming. Moreover, the reverse transcription carried out by this virus is error-prone which might result in the mutation of the transgene. Also, the packaging size of the viral capsid is moderately low which affects the payload size for recombinant protein expression (202).

The transgene expression level via random insertion is determined by its integration site in the genome, which might be transcriptionally active or repressed due to epigenetic modifications and DNA methylations (212). In addition, random integration of transgene expression construct in the genome leads to copy number variation among different clones. This will affect the testing of different variants of the transgene for better protein production. A major bottleneck in biologics production

includes high variation among clones and loss of protein productivity over time. Although initially, the protein expression is high, it reduces during prolonged culture due to the lack of stability of the transgene integrated inside the genome (212,221).

One alternative tool to achieve high transgene expressing stable clone is using site-specific recombinases, such as, Flp/*FRT*, Cre/*LoxP*, phiC31 and BxB1 integrase, which can integrate the transgene expression construct at a specific transcriptionally active genomic location (222). Another powerful strategy that has recently been developed for targeted insertion of the transgene is recombinase-mediated cassette exchange (RMCE) that uses Flp or Cre recombinase (223,224). In this strategy, a master cell line is generated by integrating a single copy of reporter gene flanked by recombinase target site (*FRT/loxP* sites) at a transcriptionally active genomic site. The transgene is introduced to the target site by a recombinase-mediated exchange of reporter gene with the transgene (figure 1.18). Hence, it provides an opportunity for site-specific integration of the transgene at a transcriptionally active genomic locus (209,223,224). It has been shown that even a single copy of transgene integration at a transcriptionally active genomic site can provide a stable and high protein-expressing clone. Thus RMCE can improve the speed and efficiency to generate a producer cell line (225). RMCE was successfully applied on CHO-K1 cells where the master cell line harboring BxB1/*att*-landing pads were targeted with large transgene up to 100kb size for the production of monoclonal antibody (145).

Figure 1.18. Recombinase-mediated cassette exchange (RMCE) between donor plasmid and genomic locus tagged by recombination target site resulting in the insertion of GFP cassette in the genome [reprinted with permission (160)].

Site-specific integration of transgene can be also achieved by using programmable endonucleases such as ZFNs, TALENs and CRISPR/Cas9. They have been successfully used for site-specific insertion of the transgene at definite loci in CHO and other mammalian cells (226-228). However, using programmable endonucleases for such an application might result in aberrant recombination events at the integration site and several off-target effects (229).

1.5. Enhancement of transgene expression by introducing a DNA element in the seamless vector

The expression of the integrated transgene depends on the chromatin structure surrounding the integration site, also known as position effect. Chromatin position effect might silence the transgene

expression when the site of the integration is in the vicinity of heterochromatin. Such a region might lead to the accumulation of highly repressive marks at the integration site by DNA methylation and histone deacetylation (230-233). Several regulatory DNA elements are being studied to overcome chromatin positional effects on transgene that includes locus control regions (LCR), ubiquitous chromatin region opening element (UCOE), insulators and scaffold/matrix attachment regions (S/MARs) (234).

1.5.1. Different DNA elements to enhance transgene expression inside the genome

- a) The LCRs are a class of DNase I hypersensitive sites mainly found upstream of the human β globin domain at the 5' noncoding region. LCRs mainly act as long-range transcriptional enhancers that can dominantly open chromatin and it has a positive effect on the expression of transgene irrespective of the chromatin integration site (235,236). However, the activity of LCRs is mostly tissue-specific, thus limiting its application for transgene expression (237). It has been found that the expression of human factor VIII and IX protein was enhanced in a liver-specific manner upon inclusion of hepatocyte-specific LCR in the transgene construct (238).
- b) Insulators are specific DNA elements that manifest a strong ability to protect transgene from chromatin positional effects. Insulators can interfere with the communication between enhancer and promoter when located in between the two. They can function independently by partitioning the genome into a discrete functional domain without disturbing the endogenous signals between enhancers and promoters (239-241). One of the well-studied insulators is chicken β globin cHS4 insulator that has been found to contribute to the stable long-term transgene expression and can also promote histone acetylation and impedes DNA methylation (233,242).
- c) Ubiquitous chromatin region opening element (UCOE) is a commonly used DNA element to protect transgene from positional effects. It consists of methylation-free CpG islands which are usually placed near the transcription start site. Inside genome, UCOE is mostly found in the vicinity of the ubiquitously expressed genes (243). It was well studied that inclusion of UCOE in the transgene construct has led to a high level of stable long-term transgene expression for over 100 generations (244-246). In addition, cell lines harboring UCOE showed to have high transcriptional activity and less clonal variation as compared to non-UCOE cell lines (247).
- d) Another effective way to overcome the positional effect is to prevent DNA methylation by using the core CpG island elements (IEs). This element is smaller than 120bp in size which makes it convenient for application as compared to other elements of size thousands of base pairs (248).

- e) Scaffold/matrix attachment regions (S/MARs) are most commonly used in CHO and plant cells to rescue transgene from chromatin positional effects (249,250). They are A+T rich sequences, size ranging from 300bp to several kilobase pairs, which anchors the chromatin to the nuclear matrix protein during the interphase (251-253). They mediate chromatin attachment to a scaffold forming DNA loop leading to participation in DNA replication, transcription and RNA processing. They mainly support chromatin opening to maintain a euchromatin state (254). Based on their properties, S/MARs can contribute to the regulation of various biological processes like shielding transgene from positional effects (221), maintaining enhancer-promoter activity (255,256) and augmentation of high transcription rates (257). It has been recently reported that the inclusion of specific human MARs with a transgene can enhance its expression and copy number of transgene insertion in CHO cells (258).

Besides the above-mentioned DNA elements, another element that has not been well studied so far with respect to transgene expression is a zinc-finger DNA binding protein known as, CCCTC-binding factor (CTCF), which is crucial for the formation of topologically associated domains (TADs) (259-261).

1.6. Aim of the project

Our laboratory has recently demonstrated that variants of phage λ Int can be used as a transgenesis tool to catalyze conservative site-specific integration into the endogenous human target sequence, termed *attH4X*, found in 1000 *Long Interspersed Elements-1 (LINE-1)* (1). With the focus to improve the λ Integrase mediated transgenesis tool in terms of efficiency and to explore its applications in the field of therapeutics and biologics production, the aim and objective of the project are designed as follows:

Aim 1: λ -Integrase mediated seamless vector transgenesis platform for therapeutic protein expression. The objectives include the development of *in vitro* and *in vivo* platforms for seamless vector generation, followed by site-specific seamless vector insertion at *attH4X* sequences into human *LINE-1* and finally as a proof of concept for therapeutic protein expression, anti-CD19 CAR seamless vector insertion and expression in human embryonic stem cells (hESCs).

Aim 2: λ -Integrase mediated seamless vector transgenesis platform for biologics production. The objectives include generation of a master cell line by randomly inserting linear landing pads carrying *attH4X* sequences in HEK293T, followed by λ -Integrase mediated targeting of landing pads with seamless target vector to screen a suitable master cell line and finally as a proof of concept, λ -Integrase mediated targeting of landing pads with full-length factor 8 (FLF8) vector for production of therapeutic factor VIII protein.

Aim 3: Enhancement of transgene expression from site-specifically targeted seamless vector carrying CTCF binding motifs.

The objectives include the addition of CTCF binding sites in seamless vectors followed by λ -Integrase mediated transgenesis of the seamless vector carrying CTCF binding sites in hESCs to understand the effect of CTCF binding site in enhancing transgene expression inside human genome.

Aim 4: λ -Integrase mediated seamless vector transgenesis in actively transcribed *LINE-1 elements*.

The objectives include design and generation of seamless vector to target *attH4X* sequence in actively transcribed *LINE-1 elements* and λ -Integrase mediated targeting of actively transcribed *LINE-1 elements* with seamless vector in hESCs

2. MATERIALS AND METHODS

2.1. Cell lines

This study used human embryonic stem cell (hESC) line “Genea047” from Genea Biocells, Australia; cancer cell line of HT1080 (fibrosarcoma) and human embryonic kidney 293T cells (HEK293T).

The Genea047 human embryonic stem cell line was derived from a donated, fully commercially consented artificial reproductive technology (ART) blastocyst, through inner cell mass (ICM) outgrowth on inactivated human feeders. The genomic analysis of the cell line shows a 46, XX karyotype and female allele pattern (262).

The HT1080 is a fibrosarcoma cell line created from tissue taken in a biopsy of a fibrosarcoma present in a 35-year-old human male (263). The cell line carries an IDH1 mutation and an activated N-ras oncogene (264).

The HEK 293T cell line is derived from the HEK 293 cell line and it expresses a mutant version of the SV40 large T antigen and confers neomycin/G418 resistance (265). Human embryonic kidney 293 cells were generated by transfection of normal human kidney cells with sheared adenovirus type 5 DNA (266).

2.2. Plasmids

Standard molecular cloning techniques were employed to generate plasmids used in this work. High fidelity *Pfu* polymerase (Thermo Fisher Scientific) was used for PCR amplifications and *E. coli* DH5 α was used for plasmid DNA amplification.

pattP4X-PGKss-EGFP-bpa-attH4X was generated by PCR amplification of the *EGFP* cassette from *pCMVssEGFP* (in-house), using the primers *PstI*_EGFP and *NotI*_attH4X_ *NheI*_bpa (includes *attH4X* sequence) (primer sequences are listed in table 3), followed by cloning into *pattP4X-PGKssPuro-bpa* (in house) between *PstI* and *NotI* sites, thus replacing the Puromycin cassette with the EGFP and *attH4X* sequence (figure 2.1).

Figure 2.1. Vector map of *pattP4X-PGKssEGFP-bpa-attH4X* consisting of PGK promoter driving EGFP gene flanked by *attP4X* and *attH4X* sequences

pattP4X-PGKss-Puro-bpa-attH4X was generated by PCR amplification of the Puromycin and poly-A cassette from *pattP4X-PGKssPuro-bpa*, using the primers Puro_fwd_PstI and Puro_rev_XbaI, followed by cloning into *pattP4X-PGKssEGFP-attH4X* between PstI and XbaI sites, thus replacing the *EGFP* with the Puromycin cassette (figure 2.2).

Figure 2.2. Vector map of *pattP4X-PGKss-Puro-bpa-attH4X* consisting of PGK promoter driving puromycin resistance gene flanked by *attP4X* and *attH4X* sequences

pattP4X-PGKss-CAR-bpa-attH4X was generated by PCR amplification of the *CAR* (chimeric antigen receptor) and poly-A cassette from *pVAX1- α CD19BBZ* (provide by Prof. Dario Campana), using the primers CAR_fwd_NcoI and CAR_rev_XbaI, followed by cloning into *pattP4X-PGKss-EGFP-bpa-attH4X* between NcoI and XbaI sites, thus replacing the *EGFP* with the *CAR* cassette (figure 2.3).

Figure 2.3. Vector map of *pattP4X-PGKss-CAR-bpa-attH4X* consisting of PGK promoter driving anti-CD19 CAR (chimeric antigen receptor) gene flanked by *attP4X* and *attH4X* sequences

pEF1-CAR-bpa was generated by PCR amplification of the CAR and poly-A cassette from *pVAX1- α CD19BBZ* (provided by Prof. Dario Campana), using the primers CAR_fwd_BamHI and CAR_rev_EcoRI, followed by cloning into *pEF1-EGFP-bpa* (in-house) between BamHI and EcoRI sites, thus replacing the EGFP with the CAR cassette (figure 2.4).

Figure 2.4. Vector map of *pEF1-CAR-bpa* consisting of EF-1 α promoter driving anti-CD19 CAR gene

pattP4X-PGKss-Puro-bpa-EF-CAR-bpa-attH4X was generated by PCR amplification of the *EF-CAR-bpa* cassette from *pEF1-CAR*, using the primers HR_NheI_EF fwd and HR_NheI_Bgh rev that adds homologous arms to PCR product. PCR products were cloned into *pattP4X-PGKssPuro-attH4X* between NheI sites by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding *EF-CAR-bpa* cassette downstream of *PGKssPuro* cassette (figure 2.5).

Figure 2.5. Vector map of *pattP4X-PGKss-Puro-bpa-EF-CAR-bpa-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; PGK promoter driving puromycin resistance gene and EF-1 α promoter driving anti-CD19 CAR gene.

pattP4X-PGKss-CAR-bpa-EF-GFP-bpa-attH4X was generated by PCR amplification of the *EF-EFGP-bpa* cassette from *pEF1-EGFP* (in-house), using the primers HR_NheI_EF fwd and HR_NheI_Bgh rev that adds homologous arms to PCR product. PCR products were cloned into *pattP4X-PGKssCAR-bpa-attH4X* between NheI sites by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding *EF-EFGP-bpa* cassette downstream of *PGKss-CAR-bpa* cassette (figure 2.6).

Figure 2.6. Vector map of *pattP4X-PGKss-CAR-bpa-EF-GFP-bpa-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; PGK promoter driving anti-CD19 CAR gene and EF-1 α promoter driving GFP gene.

pattP4X-PGKss-CAR-bpa-EF-Puro-bpa-attH4X was generated by cleaving the *GFP* cassette from *pattP4X-PGKssCAR-bpa-EF-GFP-bpa-attH4X* with the enzymes BamHI and EcoRI and replacing it with PCR amplified puromycin cassette (figure 2.7) using In-Fusion HD Cloning kit (Takara Bio USA). Puromycin cassette was amplified from *pattP4X-PGKssPuro-attH4X* using primers Puro fwd_BamHI_HR and Puro rev_EcoRI_HR.

Figure 2.9. Vector map of *pattP4X-scCTCF-UTF1-GFP-scCTCF-PGKssPuro-bpa-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; PGK promoter driving puromycin resistant gene and UTF1 promoter driving GFP gene. The *UTF1-EGFP* cassette is flanked by 14bp scrambled CTCF sites in convergent orientation.

pattP4X-CTCF-UTF1-GFP-PGKss-Puro-CTCF-attH4X was generated by PCR amplification of the Puromycin and poly A cassette from *pattP4X-PGKssPuro-bpa-attH4X*, using the primers *attP_fwd_pgk_fwd_HR* and *attH_rev_CTCF_bpa_HR* that contain 14bp CTCF sequence in the reverse primer. PCR products were cloned into *pattP4X-attH4X* (in-house) between *Bam*HI sites by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding *PGKss-Puro-bpa-CTCF* cassette in between *attP4X* and *attH4X* sequence. *UTF1-GFP* was added to this plasmid by PCR amplification of the *UTF1-EGFP* cassette from *UTF1-GFP-pattP4X-PGKssPuro* (in-house), using the primers *PGK_rev_UTF1_rev* and *attP_fwd_CTCFenhancer_fwd* that contain 14bp CTCF site in the reverse primer. PCR products were cloned into *pattP4X-PGKss-Puro-bpa-CTCF-attH4X* between *Nco*I site by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding *UTF1-EGFP-CTCF* cassette upstream of *PGKss-Puro-bpa-CTCF* cassette (figure 2.10).

Figure 2.10. Vector map of *pattP4X-CTCF-UTF1-GFP-PGKss-Puro-bpa-CTCF-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; PGK promoter driving puromycin resistant gene and UTF1 promoter driving GFP gene. The *UTF1-EGFP-PGKss-Puro* cassette is flanked by 14bp CTCF sites in convergent orientation.

pattP4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro-bpa-attH4X, *pattP4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro-bpa-attH4X* and *pattP4X-CTCF-UTF1-mNeon-PGKssPuro-bpa-CTCF-attH4X* were generated by cleaving the *GFP* cassette from all three plasmids with the enzymes *SrfI* and *BsrGI* and replacing it with PCR amplified *mNeonGreen* cassette (figure 2.11) using In-Fusion HD Cloning kit (Takara Bio USA). *mNeonGreen* cassette was amplified from *pattP4X-Dux4-mNeon-PGKssPuro-bpa-attH4X* (in house) using primers mNeon_fwd_HR and mNeon_rev_HR.

Figure 2.11. Vector maps of *pattP4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro-bpa-attH4X*, *pattP4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro-bpa-attH4X* and *pattP4X-CTCF-UTF1-mNeon-PGKssPuro-bpa-CTCF-attH4X* where the EGFP cassette was replaced by *mNeonGreen* cassette.

pPGKss-4x-attH4X-Neo-bpa was generated by cleaving out the *attP4X* sequence from *pattP4X-PGKss-Neo-bpa* (in-house) with the enzyme *EcoRI* and ligated it back to generate *pPGKss-Neo-bpa*. The plasmid is then again cleaved with the enzyme *PstI* and a synthesized DNA fragment containing *4x-attH4X* along with homologous arm was cloned at *PstI* site by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding *4X-attH4X* in between *PGKss* and *Neo* cassette (figure 2.12).

Figure 2.12. Vector map of *pPGKss-4x-attH4X-Neo-bpa* consisting of PGK promoter followed by four copies of *attH4X* sequence and neomycin resistant gene.

pPGKss-4x-attH4X-Hygro-bpa was generated by cleaving the Neomycin cassette from *pPGKss-4x-attH4X-Neo* with the enzymes BsrGI and XbaI and replacing it with PCR amplified hygromycin cassette by using In-Fusion HD Cloning kit (Takara Bio USA) (figure 2.13). Hygromycin cassette was amplified from *pattB-Hygro* (in-house) using primers Hygro_Fwd_BsrGI_HR and Hygro_Rev_XbaI_HR.

Figure 2.13. Vector map of *pPGKss-4x-attH4X-Hygro-bpa* consisting of PGK promoter followed by four copies of *attH4X* sequence and hygromycin resistant gene.

pattP4X-Puro-bpa-EF-GFP-bpa-attH4X was generated by PCR amplification of the Puromycin and poly A cassette from *pattP4X-PGKssPuro-bpa-attH4X*, using the primers Puro_XmaI_Fwd_HR and Puro_XmaI_Rev_HR that adds homologous arms to PCR product. PCR products were cloned into

pattP4X-attH4X (in-house) between *Xma*I sites by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding Puro cassette in between *attP4X* and *attH4X* sequence. *EF-GFP* was added to this plasmid by PCR amplification of the *EF-EFGP-bpa* cassette from *pEF1-EGFP* (in-house), using the primers EF_BamHI_Fwd_HR and GFP_BamHI_Rev_HR. PCR products were cloned into *pattP4X-Puro-bpa-attH4X* between BamHI site by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding EF-EGFP cassette downstream of *pattP4X-Puro-bpa* cassette (figure 2.14).

Figure 2.14. Vector map of *pattP4X-Puro-bpa-EF-GFP-bpa-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; a promoterless puromycin resistance gene and EF-1 α promoter driving EGFP gene.

pattP4X-Puro-bpa-EF-FLF8-IRES-ZsGreen-polyA-attH4X was generated by cleaving out *EF-GFP-bpa* cassette from *pattP4X-Puro-bpa-EF-GFP-bpa-attH4X* with the enzyme BamHI and replacing it with PCR amplified *EF-FLF8-IRES-ZsGreen-polyA* cassette (figure 2.15) by using In-Fusion HD Cloning kit (Takara Bio USA). The *EF-FLF8-IRES-ZsGreen* cassette was amplified from *pattP4X-EF-FLF8-IRES-ZsGreen-polyA-UTF1-Puro-attH4X* (in-house) using primers FLF8_BamHI_Fwd_HR and FLF8_BamHI_Rev_HR.

Figure 2.15. Vector map of *pattP4X-Puro-bpa-EF-FLF8-IRES-ZsGreen-polyA-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; a promoterless puromycin resistance gene and EF-1 α promoter driving full-length factor VIII gene along with ZSgreen gene linked by an IRES sequence.

pattP4X-bpa-GFP-EF-bpa-Puro-IRES-attH4X was generated by PCR amplification of the IRES cassette from *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-polyA-attH4X*, using the primers IRES_SrfI_Fwd and IRES_SrfI_Rev that adds homologous arms to PCR product. PCR products were cloned into *pattP4X-Puro-bpa-EF-GFP-bpa-attH4X* between SrfI sites by using In-Fusion HD Cloning kit (Takara Bio USA), thus adding IRES sequence upstream of Puromycin cassette. The whole cassette was PCR amplified in reverse orientation using EF_Bpa_BamHI_Fwd_HR and IRES_BamHI_Rev_HR and cloned inside *pattP4X-attH4X* (in-house) at BamHI site (figure 2.16) by using In-Fusion HD Cloning kit (Takara Bio USA).

Figure 2.16. Vector map of *pattP4X-bpa-GFP-EF-bpa-Puro-IRES-attH4X* consisting of two cassettes flanked by *attP4X* and *attH4X* sequences; a promoterless puromycin resistance gene linked to IRES sequence and EF-1 α promoter driving EGFP gene. Note: the orientation of the cassettes is in the anti-clockwise direction.

Table 3. List of Primers (5'-3')

Primer Name	Sequence
PstI_EGFP	GTACCTGCAGCCACCATGGTGAGCAAGGGCGAGGAGC
NotI_attH4X_NheI_bpa	CAGTGCGGCCGCCAACTTAATGAAATAAAGCAGGCTAGCCGACTCA CTATAGGGCGAATTGGAGC
Puro_fwd_PstI	CCAACTGCAGATGACCGAGTACAAG
Puro_rev_XbaI	GACTTCTAGATCAGGCACCGGGCTTGCGG
CAR_fwd_NcoI	CCACCATGGCCTTACCAGTGACCGC
CAR_rev_XbaI	GATATCTAGATTAGCGAGGGGGCAGGGCCTGCATG
CAR_fwd_BamHI	CGTAGGATCCCCACCATGGCCTTACCAGTGACCGC
CAR_rev_EcoRI	GACTGAATCCTTAGCGAGGGGGCAGGGCCTGCATG
HR_NheI_EF fwd	CGCCCTATAGTGAGTCGGCTAGCGGCTCCGGTGCCCCGTCAAGT
HR_NheI_Bgh rev	CTTAATGAAATAAAGCAGGCTAGCGCCGCTACAGGGCGCGTGGGGA T
Puro fwd_BamHI_HR	GTAAGCTTGGTACCGAGCTCGGATCCCCACCATGACCGAGTACAAGC CCACGG
Puro rev_EcoRI_HR	GGCTGATCAGCGAGCTAGTGAATTCTCAGGCACCGGGCTTGCGGGTC
cs_attH4X_F1	GAGTGTTTTCCAACCTTGGTTCCATT
cs_attH4X_R1	AAAACACAGCACGAGAACTTCGTGA
cs_attH4X_F2	CCTGTCTTGCTAGGTTGGGAAGT
cs_attH4X_R2	TAGAGGAATTGCTAACTAGATAAACCA
Puro Fwd 127	GAGCTGCAAGAACTCTTCCTCAC
Biotinylated Puro Fwd 127	Biotin-GAGCTGCAAGAACTCTTCCTCAC
Puro Rev 24	CACCGTGGGCTTGTAATCGGTC
PETF2-PH	PHOSPO-CATCGGTGATGTCGGCGATAT
Puro Fwd 128	CAACAGATGGAAGGCCTCCTGG
petF2RC	ATATCGCCGACATCACCGATG
bpaI	GGACAGCAAGGGGGAGGATTG
PGK_rev	CAGGTGAATATCAAATCCTCCTCG
Puro Fwd 509	CTACGAGCGGCTCGGCTTCACC
PGK Rev 187	CTGACTAGGGGAGGAGTAGAAGTGG
attP_Rev	GACTGATAGTGACCTGTTCTGTTGC
PR21	CACAGGAAACAGCTATGACCATG
PR196	CTCCGGCTCGTATGTTGTGTG
CAR Probe fwd	GGGTCTGGGACCTCTTATTTTC
CAR Probe Rev	GATAAATCCGTCCAATCCACTCAAG
4x-attH4X	CTGCAGACGCTTTATTTTCATTAAGTTGCCCCACGCTTTATTTTCATTAA GTTGCCCCACGCTTTATTTTCATTAAGTTGCCCCACGCTTTATTTTCATT AAGTTGTGTACACTGCAG
Hygro_Fwd_BsrGI_HR	TTCATTAAGTTGTGTACACCACCATGAAGAAGCCTGAG
Hygro_Rev_XbaI_HR	CTGATCAGCGAGCTCTAGACTATTCCTTTGCCCTCGGAC
Puro_XmaI_Fwd_HR	CTCCCGTCTAAGCCCGGGCCACCATGACCGAGTACAAGCCCAC
Puro_XmaI_Rev_HR	GCATGCAGATCTCCCGGGCGAGGCTAGAAGTGTGGATC
EF_BamHI_Fwd_HR	CTGCAGCCATGGGGATCCGGCTCCGGTGCCCGTCAGTG
GFP_BamHI_Rev_HR	AAGCAGGCTAGCGGATCCGCCGCTACAGGGCGCGTGGG
IRES_SrfI_Fwd	CTCCCTCCCGTCTAAGCCCGTTATTTTCCACCATATTGC
IRES_SrfI_Rev	CTCGGTCATGGTGGCCCTATTATCATCGTGTGTTTCA
EF_Bpa_BamHI_Fwd_HR	CTGCAGCCATGGGGATCCGCCGCTACAGGGCGCGTGGG
IRES_BamHI_Rev_HR	AAGCAGGCTAGCGGATCCGTTATTTTCCACCATATTGC
FLF8_BamHI_Fwd_HR	CTGCAGCCATGGGGATCCGAGTAATTCATACAAAAGGACTCGC
FLF8_BamHI_Rev_HR	AAGCAGGCTAGCGGATCCGCGAGTGAATAAATGCTTTATTTGTG
attp_fwd_pgk_fwd_HR	CTGCAGCCATGGGGATCCAATTCTACCGGTTAGGGGAGGCGC
attH_rev_bpa_HR	AAGCAGGCTAGCGGATCCCGAGGCTAGAAGTGTGGATCTCG

PGK_rev_CTCF_UTF1_rev	AGAATTGGATCCCCATGGCCACCAGGTGGCAGCCACCAGGTGGCAG GGGCGGGGCGGGGCGCGCGGGGCG
attPfwd_CTCFenhancer_fwd	GCATGCCTGCAGCCATGGCCACCAGGTGGCAGCCACCAGGTGGCAG CTAATTTTTGTATTTTTAGTAGAG
PGK_rev_ScCTCF_UTF1 rev	AGAATTGGATCCCCATGGAGGAGGCACAAGCCGACGTCCTCGCGGC GGGCGGGGCGGGGCG
attPfwd_ScCTCF_enhancer_fwd	GCATGCCTGCAGCCATGGAGGAGGCACAAGCCGACGTCCTCGCGGC CTAATTTTTGTATTTTTAGTAGAG
attH_rev_CTCF_bpa_HR	AAGCAGGCTAGCGGATCCCCACCAGGTGGCAGCCACCAGGTGGCAG CGAGGCTAGAACTAGTGGATCTCG
PGK_rev_UTF1_rev	AGAATTGGATCCCCATGGGGGCGGGGCGGGGCGCGCGGGGCG
mNeon_Fwd_HR	GCCGGACTTACTTGTACAGCTCGTCCATGCCC
mNeon_Rev_HR	GTGTGGGCGGCCCGGGCGGGCTGTGGCCCTCTCCCCATCTCCACCA TGCCGAAAAAAAAAACGCAAAG
Mneon green_fwd_578	CTATCTGAAGAACCAGCCGATGTAC
PR029	CGCTCAGGAGTTCGAGACCAG
PR024	ATTGAGTGAAGAGGTGGGGTCA
PR026	CCCACTCGCTCCTAAGGTGTAG
Mneon green_fwd_330	GAAGATGGTGCCTCCCTTACTGTTAAC
Puro Rev 102	GCTCGTAGAAGGGGAGGTTG
PGK FWD Doc	CGAGGAGGATTTGATATTCACCTGG
Hygro rev Doc	CGAAGCTGAAAGCACGAGATTCTTC
GFP fwd 599:	AACCACTACCTGAGCACCCAGTC
GFP rev_151:	GTGCAGATGAACTTCAGGGTCAG
Hygro Probe Fwd	GATTCGGAAGTGCTTGACATTGGG
Hygro Probe Rev	GGCAGTCCTCGGCCCAAAGCATCAG
EGFP Probe Fwd	CCGACCACATGAAGCAGCACG
EGFP Probe Rev	GTTCTGCTGGTAGTGGTCGGCG
Puro rev 303	GGAACCGCTCAACTCGGCCATGC
F8 rev	GGATACACCAACAGCATGAAG
EGFP fwd 114	GTCTATATCATGGCCGACAAGCAG
Biotinylated EGFP fwd	BIOTIN-CCACATGAAGCAGCAGCACTTC

2.3. Plasmid DNA purification

Plasmid DNA isolation was carried out from 5ml of overnight culture using QIAprep Spin Miniprep kit (Qiagen, GmbH) according to the manufacturer's protocol. For large-scale plasmid preparation, plasmid DNA was isolated from 100ml overnight culture using the EndoFree Plasmid Maxi Kit (Qiagen, GmbH) according to the manufacturer's protocol.

2.4. DNA and RNA quantification

Plasmid DNA, RNA and genomic DNA were quantified using the NanoDrop™ 2000 (Thermo Fisher Scientific Inc). The absorbance at 260nm (A260) by Nanodrop is used to calculate the concentration of nucleic acids in a solution. An A260 reading of 1.0 is equivalent to 50µg/ml for dsDNA and 50µg/ml for RNA.

2.5. Cell culture

HT1080 and HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) growth medium supplemented with 10% FBS, 1% L-glutamine and 100Units/ml of Penicillin and Streptomycin each (Gibco, Life Technologies) at 37°C and 5% CO₂ in humidified condition. For the selection of puromycin-resistant recombinants, puromycin (stock solution of 10mg/ml in 20 mM HEPES buffer, Gibco, Life Technologies) was added to the growth medium (1µg/ml). For HEK293T cells, selection of hygromycin-resistant recombinants was carried out by adding hygromycin (stock solution of 50mg/ml in water, Gibco, Life Technologies) to the growth medium (200µg/ml). Trypsin-EDTA (Gibco, Life Technologies) was used for detaching the adherent cells for passaging.

Human Embryonic Stem Cells (hESCs; GENE 047) were cultured at 37°C, 5% CO₂ and 5% O₂ on Collagen I-coated cell culture dishes (Biocoat, Corning) in Genea M2 Medium, (Genea Biocells, Sydney, Australia), supplemented with Penicillin and Streptomycin at 25Units/ml each (Gibco, Life Technologies). For the selection of recombinants and maintenance of targeted clones, 300ng/ml of Puromycin or 100µg/ml of Neomycin (stock solution of 50mg/ml in water, Gibco, Life Technologies) was included in the growth medium. For passaging or preparing cell suspension for reverse transfections, adherent hESCs were rinsed with 1xPBS, detached by incubating at 37°C for 3min with the passaging solution (Genea Biocells) (with a volume of 100µl per well of a 6 well plate or 1ml per 10cm dish), dislodging cells by tapping and re-suspending the cells with at least 3x volume of Neutralization solution (Genea Biocells). After counting cells in a hemocytometer (Neubauer), they were pelleted by centrifuging at 300xg for 4 min and resuspended in Genea M2 Medium to the required cell density and added dropwise to Collagen I-coated dishes.

2.6. Transfections

For transfections in HT1080, 3x10⁵ or 3x10⁶ cells were seeded per well of 6 well plates (IWAKI, Japan) or 10cm tissue culture dishes (TPP, Switzerland), respectively, in DMEM growth medium a day before transfection to obtain 70-90% confluence at the time of transfection. Transfections were carried out with Lipofectamine 2000 (Invitrogen, Life Technologies) using a DNA to Lipofectamine 2000 ratio of 1µg:3µl. For every transfection per well, DNA and Lipofectamine 2000 were incubated separately in 100µl of Opti-MEM medium (Life Technologies). The complexes were then prepared by mixing DNA and Lipofectamine 2000 reagent and incubating for 20min at room temperature. The transfection mix was added drop-wise onto the cells (under DMEM growth medium without antibiotics), and transfection was allowed to proceed for 4-6 hours before complexes were removed by fresh DMEM medium.

For transfections in HEK293T, 3x10⁵ or 3x10⁶ cells were seeded per well of 6 well plates (IWAKI, Japan) or per 10cm tissue culture dishes (TPP, Switzerland), respectively, in DMEM growth medium

a day before transfection to obtain 70-90% confluence at the time of transfection. Transfections were carried out by employing Lipofectamine 3000 (Invitrogen, Life Technologies) with DNA to Lipofectamine 3000 ratio of 1µg:3µl and DNA to P3000 ratio of 1µg:2µl. For every transfection per well, DNA along with P3000 and Lipofectamine 3000 were incubated separately in 100µl of Opti-MEM medium (Life Technologies). The complexes were then prepared by mixing DNA and Lipofectamine 3000 reagent and incubating for 10 min at room temperature. The transfection mix was added drop-wise onto the cells (under DMEM growth medium without antibiotics) and transfection was allowed to proceed for overnight before complexes were removed by fresh DMEM medium.

For transfections of hESCs, FuGENE HD Transfection Reagent (Promega) was used in a reverse transfection protocol. A DNA to FuGENE ratio of 1µg:3µl was used. Transfection mixes were prepared by first diluting DNA in 100µl of Opti-MEM and 5min later the FuGENE reagent was added to the DNA dilution, mixed and incubated for 15min at room temperature for the complexes to form. During the incubation period, hESCs were harvested (as described above) and resuspended in Genea M2 Medium (without antibiotics). The transfection complexes were added dropwise to Collagen 1-coated plates and incubated at culturing conditions for 5 min after which the harvested cells were gently pipetted to the dishes at 5×10^5 cells per well of 6 well plate and 5×10^6 cells per 10cm dish. Transfections were kept overnight under standard culture conditions for hESCs, and media containing transfection complex was replaced with fresh M2 media.

2.7. Antibiotic Selection and screening for targeted cell clones

At about 48h post-transfection, selection with the respective antibiotic in growth medium at the concentrations indicated above was initiated. Selection medium was replaced once in two days until colonies expanded to about 0.3-0.4cm in diameter. At this stage, the colonies were picked by carefully scraping patches of cells with a pipette tip and transferred to 96 well plates for clonal expansion. The clones were sequentially expanded from 96 wells to 24 wells, and subsequently in 6 well plates. Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, GmbH) as per the manufacturer's protocol.

2.8. Identification of recombination events by PCR screening

PCR was performed using GoTaq Flexi DNA polymerase (Promega) to amplify both predicted recombination junctions generated by endogenous *attH4X* targeting in HT1080 cells and hESCs. To amplify the recombinant junctions at the landing pad of HEK293T cells, PCR was performed using Q5 High-Fidelity DNA Polymerase (NEB). The primers used for all the genomic PCRs are listed in table 3. Around 200ng of genomic DNA from each recombinant clone or parental cells as a template in 50µl PCR reactions. The thermal cycling parameters used for PCRs are shown in table 4:

Table 4. PCR parameter for GoTaq Flexi DNA polymerase and Q5 High-Fidelity DNA Polymerase.

PCR parameters	GoTaq Flexi DNA polymerase (Promega)	Q5 High-Fidelity DNA Polymerase (NEB)
Initial denaturation	95°C for 5min	98°C for 30seconds
Denaturation	95°C for 1min	98°C for 10seconds
Annealing (depends on melting temperature of the primers used)	56°C for 30seconds	65°C for 30seconds
Extension	72°C for 1minute/1kb	72°C for 30seconds/1kb
Final extension	72°C for 5min	72°C for 2min
Hold	4°C	4°C

The PCR samples were analyzed by electrophoresis on 1% agarose (Seakem Agarose, Lonza, USA) gels in 1xTBE (Tris-Boric acid-EDTA buffer) containing 0.5µg/ml ethidium bromide and PCR-amplified products were compared with DNA standard markers and digitally documented under UV illumination (Quantum Vilber Lourmat, Germany). PCR-amplified products were analyzed by sequencing.

2.9. Inverse PCR

Inverse PCR and corresponding nested PCRs were performed using Long Range PCR (Qiagen GmbH). Genomic DNA from each recombinant clone or parental cells was digested with a single cutter restriction enzyme, followed by ligation using T4 DNA ligase (NEB M0202) favoring self-ligation of linearized DNA. The ligation mixture composed of 200ng of linearized genomic DNA, 1x T4 DNA ligase buffer and nuclease-free water. It was incubated at 4°C overnight. It is followed by purification of ligated DNA using PCR purification kit (Qiagen GmbH). DNA was subsequently used as a template for inverse PCR using primers in the vector in opposite orientations. Around 1–2µl of the purified inverse PCR sample was used as the template for the subsequent nested PCR. The thermal cycling parameters were: an initial denaturation step at 93°C for 3 min, 35cycles of denaturation at 93°C for 15seconds, annealing at 60°C for 30seconds (annealing temperature is dependent on melting temperature of the primers used), extension at 68°C for 5 min and a final extension step at 68°C for 5min (extension time was altered based on the predicted size of product as per 1minute/1kb). PCR-amplified products were analyzed by sequencing.

2.10. Purification of Int-h/218

Plasmid *pINT* (provided by Dr Farid John Ghadessy; p53 laboratory, A*STAR) coding for C-terminally histidine-tagged Int-h/218 (180) was transformed into *Escherichia coli* (*E. coli*) Rosetta (DE3) pLysS cells, and positive clones were screened using LB plus antibiotics (250µg/ml ampicillin and 34µg/ml chloramphenicol) plates. An overnight culture of transformed cells was inoculated in 200ml of LB containing ampicillin (250µg/ml) and chloramphenicol (34µg/ml) and incubated at 37°C until the optical density of the culture reached 0.6 at 650 nm. At this point, Int-h/218 expression was

induced with 0.5mM IPTG and the culture was grown for 6 hours at 30°C (180). All the steps below were conducted at 4°C. The cell pellet was collected and sonicated in Lysis Buffer (50mM Tris pH8.0, 2mM EDTA pH 8.0, 1M KCl, 3mM PMSF, 5mM DTT, 10mM β -mercaptoethanol, 1x protease inhibitor from Roche) on ice (267). After removing cell debris, the cell extract was mixed with a Ni-NTA slurry at 4°C overnight. The Ni-NTA slurry was loaded in a column, and the cell extract was collected as a flow-through. The slurry was washed twice with Wash Buffer (50mM Tris pH8.0, 2mM EDTA pH 8.0, 1M KCl, 3mM PMSF, 5mM DTT, 10mM β -mercaptoethanol, 20mM Imidazole, 1x protease inhibitor). The protein was first eluted in Elution Buffer 1 (50mM Tris pH8.0, 2mM EDTA pH 8.0, 1M KCl, 3mM PMSF, 5mM DTT, 10mM β -mercaptoethanol, 50mM Imidazole, 1x protease inhibitor) followed by Elution Buffer 2 (50mM Tris pH8.0, 2mM EDTA pH 8.0, 1M KCl, 3mM PMSF, 5mM DTT, 10mM β -mercaptoethanol, 100mM Imidazole, 1x protease inhibitor). Eluted fractions were screened for protein by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) and the appropriate fractions were desalted using desalting preppacked gravity-flow columns (BioRad) in Desalting Buffer (50mM Tris pH8.0, 2mM EDTA pH 8.0, 1M KCl, 3mM PMSF, 5mM DTT, 10mM β -mercaptoethanol, 10% Glycerol, 1x protease inhibitor). The desalted fractions were screened for protein by SDS-PAGE and appropriate fractions were tested for *in vitro* recombination activity, as described below.

2.11. *In vitro* recombination assays and seamless vector production

The *in vitro* recombination was carried out in a reaction mixture comprised of TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0), 600ng of supercoiled substrate plasmid, 17ng/ μ l of purified single polypeptide chain IHF (scIHF2, in house) (268), 200mM KCl and 15ng/ μ l of partially purified Int-h/218. The reaction mixture was incubated at 37°C for 2 hours, and DNA was purified using PCR Purification columns (Qiagen, GmbH). The *in vitro* recombination leads to the formation of two catenane rings. One ring is digested with an appropriate restriction enzyme(s) that will cleave in the bacterial backbone and leaves the supercoiled seamless vector intact. The entire reaction mixture was separated on 1% agarose gel in 1xTBE (Tris-Boric acid-EDTA buffer) containing 1 μ g/ml ethidium bromide at 40V overnight. The seamless vector being covalently closed migrates ahead of the linearized bacterial part of a plasmid (described in detail in the Results section). The seamless vectors were purified from agarose gels using the QIAquick Gel Extraction Kit (Qiagen, GmbH).

2.12. Quantification of gel image using ImageJ software

In the ImageJ 1.50i software, the gel image was opened from the file menu to quantify the DNA bands. A rectangular area was drawn around the first DNA band using the “Rectangular Select” tool. To mark this area as the first band, “CTRL 1” was pressed. The first rectangular box was dragged over the second DNA band and “CTRL 2” was pressed to mark this as the second band. This step

was repeated to select and mark all DNA bands that need to be quantified. Once all the bands were marked, “CTRL3” was pressed to obtain the histogram for each band. The histogram represents the intensity of each band. For each histogram, “draw line” button was selected, and a line was drawn across the top of the histogram from where it first begins to drop steeply until where it levels out again. This was followed by the selection of the “magic wand” button, and a click was made anywhere inside the histogram. The selected area would be outlined in yellow and a new window named “Results” would appear which indicates the intensity of the band as a numerical value - the brighter the band, the higher the number. This step was repeated for all the histograms and the corresponding numerical value for each band was analysed to calculate the recombination efficiency. The recombination efficiency for a reaction was calculated using the following formula: (sum of values corresponding to recombinant products/sum of all the obtained values obtained) * 100.

2.13. Transgene insertion Mapping

The transgene insertion mapping aimed to identify the genomic location of seamless vector insertion in HT1080 cells and human embryonic stem cells. The first step of insertion mapping involves a single biotinylated primer that hybridizes with a known sequence of the inserted vector. The primer will extend, be amplified and generate many molecules of the biotinylated ssDNA including the vector sequence and its point of integration into the genome. This ssDNA PCR was performed on genomic DNA from HT1080 bulk culture and hES clones (described in detail in the Results section). Biotinylated forward primer-1 was used to generate the ssDNA extending from the seamless vector to the junction of insertion of the vector into the human genome as shown in figure 2.17. PCR was performed using Bioline Taq polymerase with Tris-based buffer (Vivantis), and the reaction conditions include 95°C, 120seconds; 95°C, 5seconds; 50°C, 30seconds, 72°C, 120seconds; 50 cycles. The second step includes capture of ssDNA onto streptavidin dynabeads (M-280 Streptavidin, Thermo Fisher Scientific), which includes washing of dynabeads with 1x PCR buffer (Tris-based) followed by incubation of PCR product generated in the first step with dynabeads for 3hours on a roller at room temperature. A phosphorylated primer (pETF2-PH) was used as an adaptor and for 5'-adenylation (NEB E2610S). The ligation of the 5' adenylated oligo with captured ssDNA was catalyzed by 5' AppDNA ligase (NEB) at 65°C for an hour. Semi-nested PCR was performed on beads with forward primer-2 and reverse primer petF2RC using Taq Polymerase (Bioline) and non-Tris based buffer (Bioline). PCR reaction conditions were 95°C, 120seconds; 95°C, 5seconds; 55°C, 30seconds, 72°C, 120seconds; 34cycles. Further, nested PCR was carried with forward primer-3 and reverse primer petF2RC, and PCR products were analyzed on 1% agarose gels in 0.5x TBE buffer.

Figure 2.17. An illustration depicting the PCR strategy for mapping insertion sites of seamless vector insertion in HT1080 cells and hESCs. This includes generation of targeted ssDNA using a biotinylated primer complementary to the inserted vector sequence (step 1), capture of the ssDNA with streptavidin beads, ligation of adenylated oligonucleotide (adaptor) and PCR amplification and sequencing analysis (steps 2-6). Positions of the relevant primers used to map *attH4X* (right junction) are indicated in step 5.

2.14. Cell surface staining of anti-CD19 CAR on hESCs

For detection of anti-CD19 CAR expression on hESCs, 0.25×10^6 cells were harvested and washed twice with PBSA (0.1% BSA; 0.1% sodium azide in 1x PBS). This was followed by blocking with 2 μ l of Normal Rabbit Serum (Thermo Fischer Scientific) and staining with biotin-conjugated AffiniPure F(ab')₂ Fragment Goat-anti-Mouse IgG (Jackson ImmunoResearch, West Grove, PA, USA) at room temperature for 10 mins in the dark. Cells were washed with PBSA twice and incubated with Allophycocyanin (APC)-Streptavidin conjugate (Jackson ImmunoResearch, West Grove, PA, USA) at room temperature for 10mins in the dark. After washing cells twice with PBSA, the cells were fixed with 300 μ l of 0.5% formaldehyde. APC-stained hESCs were detected with BD LSRFortessa™ X-20 (Becton Dickinson, USA).

2.15. Flow cytometry

BD LSRFortessa™ X-20 (Becton Dickson, USA) was used to analyze and quantify GFP⁺, ZsGreen⁺ and anti-CD19 CAR⁺ cells. The cells were harvested, centrifuged and suspended in DMEM media for HEK293T cells or M2 media for hESCs. Dot plot of side scatter (SSC) versus forward scatter (FSC) was used to gate live cells to separate them from aggregated and dead cells. For gated cells, a histogram plot of GFP or ZsGreen or APC (Allophycocyanin) was constructed. The data were analyzed using FlowJo software and represented as an overlay histogram of non-transfected control population onto the transfected population that clearly showed the percentage of GFP⁺ or ZsGreen⁺ or anti-CD19 CAR-expressing population.

2.16. Southern Blot Analysis

Genomic DNA was purified using the DNeasy Blood & Tissue Kit (Qiagen, GmbH). 20µg of genomic DNA was subjected to restriction digestion using 50U of the respective enzyme in 200µl overnight at 37°C. DNA was ethanol precipitated and dissolved in 20µl TE buffer (pH 8.0). Target vectors were linearized with a single cutter restriction enzyme and diluted to 10⁷, 10⁸ copies per µl in TE buffer (pH 8.0). Digested genomic DNA samples were resolved overnight on a 0.8% agarose gel in 1x TAE (Tris-Acetate-Boric acid) buffer, with 1kb DNA marker ladder (New England Biolabs). This was followed by the following steps: (i) the agarose gel was submerged in 250mM HCl with shaking at room temperature for 10mins (ii) washed with distilled water and submerged in Denaturation Solution (0.5M NaOH, 1.5M NaCl) for 2×15 min at room temperature with gentle shaking (iii) washed with distilled water and submerged in Neutralization Solution (0.5M Tris-HCl, pH 7.5; 1.5M NaCl) for 2×15 minutes at room temperature (iv) submerged in 20x SSC buffer (3M NaCl, 0.3M sodium citrate, pH 7.0) for 10mins with shaking at room temperature. The DNA in the agarose gel was then transferred overnight onto a positively charged nylon membrane by capillary transfer method using 20x SSC buffer. After overnight transfer, the DNA was UV-crosslinked on the membrane by exposing the membrane for 1-3mins under UV light. The membrane was then probed with the respective probes, as indicated, using the DIG-High Prime DNA Labeling and Detection Starter Kit II (Roche) as per the manufacturers' protocol. The probe-target hybrids on the blots were detected by chemiluminescent assay followed by exposure to an X-ray film (Kodak MXG film, Kodak) and developed on a Kodak X-OMAT 2000 Processor.

2.17. Western blotting

Cell lysates were prepared by re-suspending cells in ice-cold RIPA buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5mM EGTA, 1%Triton X-100, 0.1% Sodium deoxycholate, 0.1% SDS and 140mM NaCl) containing 1X Protease Inhibitor Cocktail (Roche), followed by incubation on ice for 20mins. The lysates were sonicated on ice (5W, 10 × 3 s), and insoluble components were removed by

centrifugation at 12000×g for 15min. The supernatant was collected, and protein concentrations were determined with Bio-Rad Protein Assay Dye Concentrate. Proteins were separated by SDS PAGE gels (10–15% acrylamide) and then transferred onto PVDF membranes with 0.2µm pore size (Bio-Rad) using transfer buffer (25mM Tris-HCl pH 7.6, 192mM glycine, 20% methanol). Nonspecific binding was blocked by 4% skim milk in 1x PBST (1x PBS, 0.05% Tween20) for 1 hour at room temperature and then incubated overnight at 4 °C with mouse anti-human CD3ζ primary antibody (1:1000 dilution; clone 8D3; Pharmingen) in 1x PBST containing 5% BSA. The antibody binds to the signaling module CD3ζ chain of the anti-CD19 CAR protein. Amounts of β-actin protein were determined by monoclonal antibodies raised against human β-actin (1:10,000 dilution; A1978, Sigma). Blots were washed with 1x PBST and incubated for 2hours at room temperature with goat anti-mouse IgG HRP conjugated secondary antibody (1:3000, R&D systems) in 4% skim milk. After washing with 1x PBST, immunoreactive bands were detected using the Western HRP substrate (Luminata Forte, Millipore) in an infrared Imager (LAS-4000, Fuji).

2.18. Factor VIII activity Test

To test *in vitro* Factor VIII secretion from HEK293 cells, ~20000 cells were seeded in per well of 96-well plate in 100µl DMEM media without puromycin and cultured for 48 hours. The supernatants (5µl) were collected from each well of clones and the parental clone for FVIII activity test. Activity test was performed as per the kit manual (BIOPHEN FVIII:C (Chromogenic assay), HYPHEN BioMed, 221402). This assay kit is based on the ability of activated factor VIII (FVIIIa) to generate activated factor X (FXa). The secreted factor VIII (FVIII) protein from HEK293 cells is activated in the presence of calcium ions, thrombin and negatively charged membrane phospholipids. The activated factor VIII (FVIIIa) protein then binds to the activated Factor IX (FIXa). This complex then proteolytically activates factor X (FX) to activated factor Xa (FXa). The generated FXa proteolytically cleaves a specific Factor Xa chromogenic substrate (Sxa-11) and releases para-nitro aniline (pNA) chromophore, which is quantified from colour development at 405nm using a microplate reader.

The assay was performed on Corning 96-well microplate with a transparent flat bottom. The FVIIIa Enzyme stock solution (50ng/µl) was diluted with Reagent 4+ (Tris-BSA Buffer) to prepare FVIIIa Enzyme Working Solution (0.5ng/µl). This was followed by addition of 0, 2, 4, 6, 8, and 10µl of FVIIIa Enzyme working solution (0.5ng/µl) into a series of wells in the 96-well plate to prepare 0, 1, 2, 3, 4, and 5ng/well of FVIIIa Enzyme standard. Activity test for each sample was performed as triplicate. DMEM media was used as a background control for the test. Supernatants (5µl) collected from clones and parental clones were added on the plate and the volume in every well was adjusted to 20µl with Reagent 4+ (Tris-BSA Buffer). This was followed by addition of 20µl Reagent 1 (Human

Factor X) and 20µl Reagent 2 (Factor IXa, with thrombin, phospholipids and Calcium) to every well and mixed by pipetting. The plate was incubated at 37°C for 5mins. After incubation 20µl of Sxa-11 substrate was added to samples and background control wells. The color development in the plate was measured for absorbance at 405nm (A405nm) using BioTek Cytation 5 cell imaging multimode reader in kinetic mode for 5 hours at 37°C. To obtain the corresponding values of absorbance, two time points (T1 and T2) in the linear range of the plot were selected. The absorbance value of background control was subtracted from value obtained from the sample well. Absorbance values were then used to calculate the fold change of clones over parental clones for FVIII activity and the fold change was then normalized with cell viability obtained through MTT assay.

2.19. MTT assay

Cell viability was checked using MTT assay, for which ~20000 cells were seeded per well of 96-well plate in 100µl DMEM media without puromycin and grown for 48 hours. 10µl of MTT reagent (5mg/ml, dissolved in PBS) was added to each well of seeded cells. The plate was incubated at 37°C for 30 mins. After incubation, media containing the MTT reagent was carefully aspirated from each well and 100µl of DMSO was added in the centre of each well. The plate was covered with foil and shaken on an orbital shaker for 5mins. The plate was read for absorbance at 570nm and data was recorded to calculate cell viability for each well.

2.20. *In vivo* generation of the seamless vector

Substrate plasmid was transformed into competent *Escherichia coli* (*E. coli*) LZ41 cells (269) and positive clones were screened using LB plus antibiotics (250µg/ml ampicillin) plates incubated at 30°C. An overnight culture of *E. coli* LZ41 cells transformed with substrate plasmid was sub-cultured in 100ml of LB containing ampicillin (100µg/ml) and incubated at 30°C until optical density reached 0.5 at 600nm. At this point, the culture was induced by incubating it at 42°C for 30mins to activate λ -Integrase expression. After induction, the culture was incubated at room temperature for 3hours to facilitate *in vivo* intramolecular recombination. The cell pellet was collected at room temperature and plasmid was extracted from the cells using Plasmid Maxi Kit (Qiagen, GmbH) as per the manufacturer's protocol. The isolated plasmid was then digested overnight with a suitable restriction enzyme (single cutter on the bacterial sequence of plasmid) and T5 exonuclease (NEB M0363) at 37°C. The seamless vector was purified from the digestion mixture using Phenol-Chloroform extraction and ethanol precipitation of DNA. The obtained pellet for seamless vector was then dissolved in appropriate buffer, quantified and were analyzed by electrophoresis on 1% agarose (Seakem Agarose, Lonza, USA) gels in 1xTBE (Tris-Boric acid-EDTA buffer) without ethidium bromide.

3. RESULTS

3.1. λ -Integrase mediated seamless vector transgenesis platform for therapeutic protein expression.

3.1.1. Strategy to generate seamless DNA target vectors via *in vitro* intramolecular recombination for subsequent genome editing

The integration system of phage λ has been recently developed as a novel transgenesis tool for human genome engineering (1). This tool comprises of a highly active, co-factor independent λ integrase variant, Int-C3, and two DNA sequences that act as recombination partners for intermolecular recombination. The two sequences, known as attachment (*att*) sites, includes: *attP4X* of 241bp that is placed on the target vector and *attH4X* of 18bp that is present in open reading frame 1 (ORF1) of a small subset of human Long INterspersed Elements-1 (*LINE-1*) dispersed at 933 different genomic loci. The λ Int-mediated transgenesis tool involves intermolecular recombination by λ integrase variant, Int-C3, between genomic *attH4X* (18bp) and *attP4X* (241bp) on circular target vectors. It results in the integration of the entire bacterial plasmid into *LINE-1* element flanked by hybrid *attL4X* on the left junction and *attR4X* sites on the right junction (figure 3.1 a). Chandra *et al.* also reported that the *attP4X* sequence on target vector can be functionally replaced by the shorter 120bp *attL4X* hybrid recombination site for Int-mediated recombination with *attH4X* in *LINE-1* (figure 3.1 b) (1).

Figure 3.1. λ -Integrase mediated transgenesis tool for human genome engineering (a) A schematic diagram showing λ -Int-mediated recombination between genomic *attH4X* (located in *LINE-1* elements) and *attP4X* (present on a plasmid target vector). The recombination leads to the site-specific integration of the entire plasmid into the human *LINE-1* elements, flanked by *attL4X* (left junction) and *attR4X* (right junction) [reprinted with permission (181)]. (b) A schematic diagram showing λ -Int-mediated recombination between genomic *attH4X* and *attL4X* on target vector (replaced over *attP4X*) that leads to the site-specific integration of the target vector into the human *LINE-1* elements, flanked by *attL4X* (left junction) and *attH4X* (right junction).

However, one of the major problems of this tool is the requirement of a circular target vector for Int-mediated recombination into genomic *attH4X*. Hence, the target vector needs to be produced in bacterial cells that require the presence of a bacterial backbone in the target vector construct for its amplification

and selection in bacteria cells. The bacterial backbone consisting of antibiotic-resistant genes, an origin of replication, potential cryptic expression signals, and unmethylated CpG motifs (182) also get inserted into the human genome with the target vector. It might lead to silencing of the transgene and activation of immune responses (191). Moreover, the bacterial sequences are unnecessary with respect to gene transfer and can be associated with safety issues in clinical applications. Thus, to overcome this issue and to improve the λ -Integrase mediated transgenesis tool with respect to efficiency and safety concerns, the aim was to eliminate bacterial sequences by generating seamless vectors for targeting of endogenous *attH4X* sites in *LINE-1* elements. The strategy involves (i) to generate seamless vectors via *in vitro* *attH4X* x *attP4X* intramolecular recombination on a supercoiled substrate plasmid (8) to target the obtained seamless vector carrying *attL4X* at genomic *attH4X* sites inside human cells for transgenesis (figure 3.2). The substrate plasmid carries direct repeats of *attP4X* and *attH4X* which can undergo intramolecular recombination in the presence of Int leading to the formation of two interlinked DNA circles. This seamless vector can be isolated from interlinked DNA circles by endonucleolytic digestion of the undesired bacterial DNA circle. This will result in a linearized bacterial DNA molecule and supercoiled seamless vector (figure 3.2). One of the distinct features of the seamless vector carrying *attL4X* produced from this *in vitro* recombination is that it can recombine with the *attH4X* present in *LINE-1* of the human genome (1) via Int-C3 mediated recombination for insertion of transgene cassette.

Figure 3.2. An illustration depicting *in vitro* intra-molecular *attH4X* x *attP4X* recombination of a substrate plasmid, resulting in the formation of a seamless vector carrying the *attL4X* sequence after releasing the undesired bacterial DNA ring by endonucleolytic digest. The seamless vector is subsequently targeted at genomic *attH4X*, leading to the insertion of payload through a second λ -Int-mediated recombination event [reprinted with permission (181)].

3.1.2. Partial purification of λ Integrase and generation of seamless vector

The production of seamless vectors via *in vitro* *attH4X* x *attP4X* intramolecular recombination requires purified Int protein. Hence, C-terminally histidine-tagged Int-h/218 or Int-C3 were overexpressed in *Escherichia coli* (*E. coli*) Rosetta (DE3) pLysS cells. IPTG-induced *E. coli* cultures showed overexpression of intracellular C-terminally histidine-tagged Int-h/218 and Int-C3 proteins as compared to uninduced culture. The induced cell extract of *E. coli* (Rosetta (DE3) pLysS cells) contains several other bacterial proteins along with C-terminally histidine-tagged Int-h/218 and Int-C3 protein at ~40 kDa (as indicated by the arrows, figure 3.3).

Figure 3.3. SDS-PAGE gel image depicting *E. coli* cell extract of an induced culture overexpressing C-terminally histidine-tagged Int-h/218 (left) and Int-C3 (right) at ~40kDa (pointed by arrows) upon addition of IPTG as compared to uninduced cell extract. L, PageRuler Prestained protein molecular weight marker (10kDa-250kDa).

To test whether the cell extracts contained highly active Int-h/218 or Int-C3 protein, they were used to catalyze *in vitro* intramolecular recombination between direct repeats of *attH4X* and *attP4X* present on a supercoiled substrate plasmid. This will result in the formation of two interlinked DNA circles. When the DNA circles were linearized with an appropriate restriction enzyme, they separate from each other, such that *attL4X* will join the payload DNA and *attR4X* the undesired bacterial backbone sequence from the original plasmid (figure 3.4 a). To generate seamless DNA, the substrate plasmid carrying *attH4X* and *attP4X* was mixed with *E. coli* cell extract containing Int-h/218 in TE buffer (pH8.0) and incubated at 30°C for 1.5 hr. No extra salt was added in the reaction mixture as the lysis buffer used to prepare cell extracts contained 1M KCl that would participate in the recombination reaction. The activity of Int proteins in cell extracts for *in vitro* recombination was

screened in the absence and presence of purified scIHF2 (single polypeptide chain IHF) [in house] (268) which was added in the reaction mix. Integration host factor (IHF) is a heterodimeric *E. coli* protein that acts as an essential cofactor for λ -integrase mediated integrative and excisive recombination pathway (270). Its ability to introduce DNA bending, allows it to serve as a crucial architectural protein for recombinogenic nucleoprotein complexes assembly. The main function of IHF is to aid the generation of high-precision DNA loops that permit integrase monomers to interact with two separate DNA sites simultaneously (271). scIHF2 (single polypeptide chain IHF) is a modified version of wild-type IHF composed of the alpha subunit of IHF inserted into the beta subunit through a novel protein engineering approach. It exhibits similar properties to wild-type IHF in term of DNA binding and bending (268). By quantifying figure 3.4 b using ImageJ software, it was found that scIHF2 substantially enhances the *in vitro* recombination between *attH4X* and *attP4X* in the presence 5 μ l of cell extract containing Int-h/218 resulting in ~45% recombination of the original substrate.

Figure 3.4. Generation of the seamless vector in the presence/absence of scIHF2. a) An illustration depicting *in vitro* intra-molecular *attH4X* x *attP4X* recombination of a substrate plasmid, resulting in the formation of a linearized seamless vector and bacterial DNA after endonucleolytic digestion of interlinked DNA circles (b) Agarose gel electrophoresis depicting *in vitro* intra-molecular Int-h/218 mediated *attH4X* x *attP4X* recombination products obtained with different volumes of cell extracts containing Int protein in the presence and absence of scIHF2. After recombination in the presence of Int and scIHF2, restriction digestion results in linearized bacterial DNA and the seamless vector at 4.9kb and 3.4kb, respectively. In the absence of integrase (-Int), restriction digestion resulted in linearized substrate plasmid at 2.5kb and 5.9kb bands. Partial restriction digestion of substrate plasmid resulted in linearized plasmid at 8.3kb.

To further improve the efficiency of seamless vector generation, the *in vitro* recombination was performed with increasing amounts of cell extract from 5 μ l to 10 μ l and incubating at 37°C instead of 30°C for 1.5 hrs. In this case, only the undesired bacterial DNA molecule was linearized with an appropriate restriction enzyme. The seamless vector, being negatively supercoiled, migrated faster than linearized bacterial backbones upon separating through agarose gel electrophoresis. By quantifying figure 3.5 a, b using ImageJ software, it was found that the recombination efficiency between *attH4X* and *attP4X* was enhanced to ~75% with larger volumes of added cell extract for Int-h/218 in the presence of scIHF2. However, with a larger volume of added cell extract having Int-C3

the recombination efficiency was increased to ~56% in the presence of scIHF2. Moreover, upon storage over a long period, the cell extract for Int-C3 showed slightly lesser activity as compared to the cell extract for Int-h/218. Hence, Int-h/218 was further used for seamless vector generation.

Figure 3.5. Representative analysis of an *in vitro* intra-molecular Int-h/218 (a) and Int-C3 (b) mediated *attH4X* x *attP4X* recombination with different volume of cell extract having Int protein. Supercoiled seamless vector (SC-SV) and open circular seamless vector (OC-SV) obtained after λ -Int-mediated recombination were resolved on 1% agarose gel in the presence of ethidium bromide. This resulted in ~1.2kb and ~2kb bands for SC-SV and OC-SV, respectively whereas the restriction digested, linearized bacterial catenane DNA and the unrecombined substrate plasmid were observed at ~3kb and ~5.0kb, respectively. NC: negative control reaction without integrase protein and scIHF2.

Int-h/218 was further purified to improve activity and shelf life. The Int-h/218 (his-tagged at C terminal end) was overexpressed in *E. coli* culture (as indicated by an arrow in figure 3.6 a) and purified using Ni-NTA resin. The protein was eluted from Ni-NTA resin at 50mM and 100mM imidazole. After analysing the eluted fractions on SDS-PAGE, it was noticed that the eluted fractions E2 to E8 were having a high amount of Int protein (figure 3.6 b). These fractions were processed further by desalting using a desalting column (figure 3.6 c). The desalted fractions were collected and desalted fractions 1 and 2 were estimated to have ~3mg/ml of protein. Hence, around 5-6mg of partially (about 90%) purified variant Int-h/218 were obtained from 200ml of *E. coli* culture, which was stored at -20°C.

Figure 3.6. Partial purification of C-terminally histidine-tagged Int-h/218. a) SDS-PAGE gel image depicting *E. coli* cell extract of an induced culture expressing C-terminally histidine-tagged Int-h/218 at ~40kDa upon addition of IPTG as compared to uninduced cell extract. L, PageRuler Prestained protein ladder (10kDa-250kDa). b) SDS-PAGE gel image depicting induced cell extract (In), flow-through (F), wash fractions (W1, W2) of Ni-NTA column and eluted fractions from Ni-NTA column of Int protein with 50mM imidazole (E1-E5) and 100mM imidazole (E6-E10). c) SDS-PAGE gel image depicting different desalted fractions of partially purified C-terminally histidine-tagged Int-h/218 [reprinted with permission (181)]. E, eluted fraction from the desalting column; 1,2,3,4, different desalted fractions of partially purified Int-h/218.

The desalted fraction of partially purified Int-h/218 was analysed for recombination activity by catalysing *in vitro* recombination of a 4.9kb substrate plasmid carrying direct repeats of *attH4X* and *attP4X* and a puromycin expression cassette as payload (figure 3.7 a). The recombination activity was analysed with different concentrations of Int-h/218 protein and potassium chloride (KCl). It was observed that the partially purified Int-h/218 was functional for recombination activity in the presence of high salt concentration (~150mM KCl). The partially purified Int-h/218 was functional in the presence of 100-200mM KCl, beyond which high salt concentrations inhibited the recombination activity (figure 3.7 b). This result was consistent with the seamless vector production employing crude cell extract, where the reaction mixture also contained 150mM KCl upon using 10µl of cell extract. In addition, by quantifying figure 3.7 b using ImageJ software, it was found that a recombination efficiency of ~45% for *in vitro* recombination can be achieved with only a small amount of partially purified Int-h/218. Hence, the procedure for seamless vector generation was finally optimized to 150mM KCl along with 15ng/µl of partially purified Int-h/218 and 17ng/µl of purified scIHF2 (268) in the reaction buffer. The seamless vector was purified from the agarose gel (figure 3.7 b) and analyzed by agarose gel electrophoresis that showed expected supercoiled and open circular forms (figure 3.7 c).

Figure 3.7. Recombination activity of partially purified Int-h/218 (a) An illustration depicting *in vitro* intra-molecular recombination of a 5kb substrate plasmid comprises of a puromycin expression cassette flanked by direct repeats of *attH4X* and *attP4X* sequences, that resulted in the formation of a supercoiled seamless vector and linear bacterial DNA after endonucleolytic digestion of interlinked DNA circles. (b) Representative analysis of an *in vitro* intra-molecular *attH4X* x *attP4X* recombination with reaction buffers containing increasing KCl concentrations as indicated (left). Supercoiled seamless vector (SC-SV) and open circular seamless vector (OC-SV) obtained after λ -Int-mediated recombination were resolved via agarose gel electrophoresis in the presence of ethidium bromide. This resulted in DNA bands for SC-SV and OC-SV, which migrated at ~1.2kb and ~2.2kb, respectively, whereas the linearized second catenane DNA and the unrecombined substrate plasmid migrated at ~3kb and ~5.0kb, respectively, as indicated. NC: negative control reaction without integrase protein and sciHF2. (c) Agarose gel electrophoresis of purified seamless vector showing supercoiled seamless vector (SC-SV) and open circular seamless vector (OC-SV) [reprinted with permission (181)].

3.1.3. *In vitro* recombination by Int-h/218 with linear substrate plasmid

To examine recombination efficiency of partially purified Int-h/218 on linear substrate plasmid, a 4.9kb substrate plasmid carrying direct repeats of *attH4X* and *attP4X* and a puromycin expression cassette as the payload was linearized before carrying out intramolecular recombination (figure 3.8 a). The recombination was performed in the presence of sciHF2 in different concentrations of KCl. Intramolecular recombination on the linearized substrate plasmid (5kb) would result in the linearized bacterial backbone and covalently closed seamless vector (figure 3.8 a). The recombinant products were resolved via agarose gel electrophoresis in the presence of ethidium bromide. It showed linearized substrate plasmid (4.9kb), linearized bacterial backbone (3.1kb) and open circular seamless vector (~2kb) (figure 3.8 b). In addition to these products, a DNA smear was also obtained as recombination product which might correspond to of positively supercoiled seamless vector (figure 3.8 b). The positively supercoiled seamless vector appeared as a smear during agarose gel electrophoresis due to the presence of ethidium bromide in the gel, which intercalates into DNA, reduces twist of the DNA and thus introduces positive supercoiling to the covalently closed seamless vector. Integrase mediated recombination with linearized plasmid resulted in the formation of positively supercoiled seamless vector as a contrast to the negatively supercoiled seamless vector obtained by the conventional method of seamless vector generation. To further confirm, the extracted DNA from the presumed positively supercoiled SV DNA smear was digested with a restriction

enzyme and, as expected, a DNA of ~1.8kb was obtained as per the predicted linearized seamless vector (figure 3.8 c). By quantifying figure 3.8 b using ImageJ software, the recombination efficiency of Int-h/218 mediated recombination on the linearized plasmid at 150mM KCl was found to be ~78%, that suggests more efficient recombination as compared to supercoiled substrates.

Figure 3.8. Int-h/218 mediated intramolecular recombination on a linear substrate plasmid (a) Representative analysis of an *in vitro* intra-molecular Int-h/218 mediated *attH4X* x *attP4X* recombination in a linearized plasmid that resulted in linear bacterial backbone and covalently closed seamless vector (b) Relaxed seamless vector (relaxed-SV) and open circular seamless vector (OC-SV) obtained after λ -Int-mediated recombination were resolved via agarose gel electrophoresis in the presence of ethidium bromide as shown in the left panel. This resulted in ~1.3kb smear for relaxed-SV and ~2kb band for OC-SV, whereas the restriction digested, linearized bacterial DNA and the substrate plasmid were obtained at ~3kb and ~5.0kb, respectively. (c) The digested (1.8kb) and relaxed (2.1kb) seamless vectors analysed by agarose gel electrophoresis are shown at the right. NC: negative control reaction without integrase protein and scIHF2.

3.1.4. *In vitro* recombination by Int-h/218 with *attH4X* x *attP4X* vs *attL4X* x *attR4X*

To determine the most efficient recombination partner pairs for Int-h/218 between *attH4X* x *attP4X* and *attL4X* x *attR4X*, two plasmids, p λ IR carrying wild-type *attP* and *attB* sequences and p λ ER carrying wild-type *attL* and *attR* sequences were employed for *in vitro* recombination (178) (Figure 3.9 a, b). The *in vitro* recombination was carried with both plasmids by partially purified Int-h/218 in the presence/absence of scIHF2. In the presence of Int and scIHF2, recombination efficiency of *attP* and *attB* was found to be about three-fold higher than *attL* x *attR* (upon quantifying figure 3.9 c using ImageJ software) suggesting the *attH4X* and *attP4X* are the most efficient recombination partners.

Figure 3.9. *In vitro* intramolecular recombination by Int-h/218 with attH4X x attP4X vs attL4X x attR4X. An illustration depicting *in vitro* intramolecular recombination of pλIR (a) and pλER (b), resulting in the formation of a linearized seamless vector and bacterial DNA after endonucleolytic digestion of DNA catenane ring. (c) Agarose gel depicting *in vitro* intra-molecular Int-h/218 mediated recombination in pλIR and pλER in the presence/absence of scIHF2. After recombination in the presence of Int, restriction digestion resulted in linearized bacterial DNA and the seamless vector at 4.9kb and 3.4kb respectively (marked with *). In the absence of Int (-Int), restriction digestion resulted in linearized substrate plasmid at 2.5kb and 5.9kb bands. Partial restriction digestion of substrate plasmid resulted in linearized plasmid at 8.3kb. P, supercoiled plasmids pλIR and pλER.

3.1.5. λ-Integrase mediated *in vivo* production of the seamless vector in *E. coli* strain LZ41

3.1.5.1. Strategy for the generation of the seamless vector *in vivo*

To further expand the λ-mediated seamless vector production to a large scale, the aim was to scale up the production of the seamless vector using an *in vivo* model. With this respect, two previously established *E. coli* strains, LZ41 and LZ54, were utilized (269,272-274). These strains were generated by lysogenizing wild type *E. coli* strain with a λ-phage that encodes wild type integrase and a thermolabile repressor (cI857) controlling integrase expression from the *pL promoter*. These strains

carry mutations that inactivated the *xis*, *red* and *P* gene which are required for excision, general recombination, and phage replication respectively. Besides these mutations, the strains also harbour point mutation in *topA* gene that reduces topoisomerase I activity by 99% (269,272-274). In addition, strain LZ41 carries mutation (*parC*^{K84}) in *parC* gene and strain LZ54 carries mutation (*gyrA*^{L83}) in *gyrA* gene that renders the topoisomerase IV subunits and DNA gyrase highly resistant to norfloxacin, respectively. Norfloxacin is known to block DNA decatenation in wild-type *E. coli* (269,272-274). For minicircle generation, these strains were transformed with a parental plasmid carrying wild type *attP* and *attB* sequences. Upon induction of the culture at 42°C, the thermolabile repressor (cI857) was inactivated that led to the synthesis of integrase protein (269). Since integrase protein was non-functional at 42°C (269), hence, the culture temperature was maintained at 30°C for accomplishing integrase mediated intramolecular recombination for minicircle generation.

The procedure to generate seamless vector *in vivo* using these strains includes the transformation of substrate plasmid into the *E. coli* strains, followed by induction of the culture for integrase expression (274). Upon synthesis of integrase, it would act on directly repeated *attH4X* and *attP4X* present on a substrate plasmid facilitating intramolecular recombination that would result in the formation of two interlinked DNA circles comprising of *attL4X* carrying payload and *attR4X* carrying bacterial backbone. The interlinked DNA circles will be decatenated by topoisomerase IV inside *E. coli* cells resulting in supercoiled seamless vector and bacterial backbone (figure 3.10). The DNA circles were purified from the *E. coli* strains and digested with an appropriate restriction enzyme to linearize the bacterial backbone and unrecombined substrate plasmid. This was followed by exonuclease treatment to digest the linear DNA products. Hence, only supercoiled and open circular forms of the seamless vector will be left behind which could be further purified from the reaction mixture by phenol-chloroform extraction and ethanol precipitation.

Figure 3.10. Experimental strategy for *in vivo* generation of a seamless vector in *E. coli* strains.

3.1.5.2. Optimization of culture condition of *E. coli* strain LZ41 for *in vivo* production of seamless vector

Both *E. coli* strains LZ41 and LZ54 were initially tested for efficient generation of the seamless vector. For testing, a substrate plasmid, *pattP4X-Puro-EF-GFP-attH4X*, of size 6.9kb (figure 3.11 a) was transformed in both the strains and a positive clone was selected for *in vivo* seamless vector production. In this case, the exonuclease treatment was excluded to compare the efficiency of seamless vector production between the strains. As a negative control (also termed as uninduced culture) 5ml of culture was excluded from the induction process. The culture condition includes induction at 42°C for 30mins, followed by incubating the culture at 30°C for 1hr with shaking. The plasmid was isolated from the strains, digested with an appropriate enzyme and resolved via agarose gel electrophoresis. In induced culture, intramolecular recombination and digestion with restriction enzyme resulted in linear substrate plasmid (6.9kb), linear bacterial backbone (3.1kb), and 3.8kb supercoiled seamless vector (positioned near 2kb) (figure 3.11 b). Besides these band, an extra band at ~11kb appeared for the induced digested lane (figure 3.11 b). Upon extracting the ~11kb band and digesting it with a suitable enzyme as shown in figure 3.11 c (left panel), three bands (3.3kb, 3.6kb and 3.8kb) were obtained via agarose gel electrophoresis of the digested product (figure 3.11 c, right panel). These bands suggest that the ~11kb band might correspond to a linear recombinant product that was formed due to intermolecular recombination between *attL4X* on seamless vector and *attH4X* on substrate plasmid (figure 3.11 c). In uninduced culture, there is only the linear substrate plasmid (6.9kb) and a small amount of linear bacterial backbone (3.1kb) suggesting leaky expression of integrase in both the strains (figure 3.11 b). Upon quantifying figure 3.11 b using ImageJ software, it was found that the strain LZ41 showed higher recombination efficiency of ~60% with respect to seamless vector production as compared to strain LZ54 that showed only ~35% of recombination efficiency. A similar outcome was observed when both the strains were tested with a large size plasmid, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*, of size 14kb. Hence, strain LZ41 was selected for further optimization of culture condition to generate seamless vector.

Figure 3.11. Generation of the seamless vector *in vivo* from *E. coli* strains LZ41 and LZ54. (a) An illustration depicting *in vivo* intramolecular *attH4X* x *attP4X* recombination of a substrate plasmid, resulting in the formation of two interlinked DNA circles. Upon decatenation by topoisomerase IV inside *E. coli* cells, a supercoiled seamless vector carrying *attL4X* and a supercoiled bacterial DNA ring carrying the *attR4X* were generated. (b) Agarose gel depicting *in vivo* intramolecular λ -Int mediated *attH4X* x *attP4X* recombination products obtained from induced and uninduced cultures of *E. coli* strains LZ41 and LZ54 transformed with a substrate plasmid. In the induced culture, after λ -Int mediated recombination, the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (6.9kb), linearized bacterial backbone (3.1kb), supercoiled (SC) and open circular (OC) seamless vector (SV; 3.8kb) and linear recombinant product of seamless vector and substrate plasmid (11kb). Whereas the undigested plasmid from induced culture showed supercoiled substrate plasmid, bacterial backbone, seamless vector and open circular seamless vector. In the uninduced culture (absence of recombination), the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (6.9kb), whereas undigested plasmid showed different conformations of substrate plasmid. L: 1kb ladder; Undig: plasmid isolated from *E. coli* strain LZ41 and LZ54 induced and uninduced culture; Dig: restriction enzyme *ClaI* digested plasmid isolated from *E. coli* strain LZ41 and LZ54 induced and uninduced culture. (c) An illustration depicting *in vivo* intermolecular recombination between *attL4X* on seamless vector and *attH4X* on substrate plasmid, resulting in the formation of a linear recombinant plasmid of size 10.7kb upon *ClaI* digestion. The digestion of the linearized recombinant plasmid with *BsrGI* resulted in the generation of three bands of size 3.3kb, 3.8kb and 3.6kb. The digested linear products of size 3.3kb, 3.8kb and 3.6kb are depicted in the 3rd well of the agarose gel. L: 1kb ladder; Undig: linear recombinant plasmid; Dig: *BsrGI* digested linear recombinant plasmid.

Different culture conditions were tested to increase the efficiency of seamless vector production in *E. coli* strain LZ41. For example, the induction time was increased to 40mins followed by incubation at 37°C for 1hr with shaking that resulted in the reduction of efficiency of seamless vector generation, suggesting, *in vivo* activity of integrase is not favoured at 37°C (data not shown). In another condition, the culture was induced at 42°C for 30mins followed by incubation of culture at 30°C or room temperature for 1hr, 2 hrs and 3 hrs with shaking. Upon quantifying figure 3.12 using ImageJ software, it was found that the culture kept at room temperature for 3 hrs showed higher efficiency of ~85% for seamless vector generation as compared to other cultures (figure 3.12).

Figure 3.12. Optimization of culture condition in *E. coli* strains LZ41 for *in vivo* seamless vector production. Representative analysis of an *in vivo* intra-molecular λ -Int mediated *attH4X* x *attP4X* recombination products obtained from induced and uninduced cultures of *E. coli* strains LZ41 transformed with a substrate plasmid, incubated at 30°C and room temperature for 1hr, 2 hrs and 3 hrs with shaking. In the induced culture, after λ -Int mediated recombination, the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (6.9kb), linearized bacterial backbone (3.1kb), supercoiled (SC) and open circular (OC) seamless vector (SV; 3.8kb) and linear recombinant product of seamless vector and substrate plasmid (11kb). In the uninduced culture (absence of recombination), the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (6.9kb). L: 1kb ladder; UN: plasmid isolated from *E. coli* strain LZ41 uninduced culture; 30°C: restriction enzyme *ClaI* digested plasmid isolated from *E. coli* strain LZ41 induced culture incubated at 30°C for 1hour, 2 hours and 3 hours; RT: restriction enzyme *ClaI* digested plasmid isolated from *E. coli* strain LZ41 induced culture incubated at room temperature for 1hour, 2 hours and 3 hours.

To further optimize the culture conditions for strain LZ41, the induced culture was incubated at room temperature for 3hrs and 6hrs with and without shaking. By quantifying figure 3.13 using ImageJ software, it was found that all the cultures showed an equal recombination efficiency of 50-60% for seamless vector production was observed (figure 3.13 a). Hence, the culture induction was validated at 42°C for 30mins followed by incubation at room temperature for 3 hrs without shaking for *in vivo* generation of a seamless vector in strain LZ41. Using this optimized culture condition, *in vivo* seamless vector was generated with a large size plasmid, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*, which also showed an efficiency of ~40% for seamless vector generation (figure 3.13 b).

Figure 3.13. (a) Optimization of culture condition in *E. coli* strains LZ41 for *in vivo* seamless vector production. Representative analysis of an *in vivo* intra-molecular λ -Int mediated *attH4X* x *attP4X* recombination products obtained from induced and uninduced cultures of *E. coli* strains LZ41 transformed with a substrate plasmid, incubated at room temperature for 3 hrs and 6 hrs with shaking and without shaking. In the induced culture, after λ -Int mediated recombination, the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (6.9kb), linearized bacterial backbone (3.1kb), supercoiled (SC) and open circular (OC) seamless vector (SV; 3.8kb) and linear recombinant product of seamless vector and substrate plasmid (11kb). In the uninduced culture (absence of recombination), the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (6.9kb). L: 1kb ladder; UN: plasmid isolated from *E. coli* strain LZ41 uninduced culture; NSH: restriction enzyme *ClaI* digested plasmid isolated from *E. coli* strain LZ41 induced culture incubated at room temperature without shaking; SH restriction enzyme *ClaI* digested plasmid isolated from *E. coli* strain LZ41 induced culture incubated at room temperature with shaking. (b) Agarose gel depicting *in vivo* intra-molecular λ -Int mediated *attH4X* x *attP4X* recombination products obtained from the induced and uninduced culture of *E. coli* strain LZ41 transformed with a substrate plasmid, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*. In the induced culture, after λ -Int mediated recombination, the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (14.1kb), linearized bacterial backbone (3.1kb), supercoiled (SC) and open circular (OC) seamless vector (SV; 11kb). Whereas the undigested plasmid from induced culture showed supercoiled substrate plasmid, bacterial backbone, seamless vector and open circular seamless vector. In the uninduced culture (absence of recombination), the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (7.2kb), whereas undigested plasmid showed different conformations of substrate plasmid. L: 1kb ladder; UN: plasmid isolated from *E. coli* strain LZ41 uninduced culture; Digested: restriction enzyme *ClaI* digested plasmid isolated from *E. coli* strain LZ41 induced and uninduced culture; Undigested: plasmid isolated from *E. coli* strain LZ41 induced and uninduced culture.

3.1.5.3. Large scale *in vivo* production of the seamless vector from *E. coli* strain LZ41

Upon successful optimization, the next focus of the study was to scale-up the *in vivo* seamless vector production. A substrate plasmid, *pattP4X-EF-TCR-attH4X*, of size 7.2kb encoding a therapeutic protein T-cell Receptor (TCR) was utilized for seamless vector production. To carry out, 100ml of culture was induced at 42°C for 30mins followed by incubation at room temperature for 3 hrs. Plasmid DNA was isolated and a fraction was first analysed for recombination efficiency by digesting it with a restriction enzyme. Restriction digestion of the isolated plasmid resulted in linear substrate plasmid (7.2kb), linear bacterial backbone (3.1kb), 4.1kb supercoiled seamless vector (positioned near 2.7kb) and linear recombinant product of the seamless vector and substrate plasmid at ~11kb. Whereas in the uninduced culture, there is only a band of linear substrate plasmid at 7.2kb (figure 3.14 a). The supercoiled seamless vector was analysed by agarose gel electrophoresis after exonuclease treatment (figure 3.14 b). The final yield of the seamless vector was found to be ~20 μ g from a total plasmid of

40µg indicating the efficiency of 50% for *in vivo* intramolecular recombination. This further demonstrated the utility of the developed platform for large scale production of seamless vector for downstream applications.

Figure 3.14. Generation of the seamless vector *in vivo* from *E. coli* strains LZ41 on large scale. Agarose gel depicting *in vivo* intra-molecular λ -Int mediated *attH4X* x *attP4X* recombination products obtained from the induced and uninduced culture of *E. coli* strain LZ41 transformed with a substrate plasmid. In the induced culture, after λ -Int mediated recombination, the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (7.2kb), linearized bacterial backbone (3.1kb), supercoiled (SC) and open circular (OC) seamless vector (SV; 4.1kb) and linear recombinant product of seamless vector and substrate plasmid (11kb). Whereas the undigested plasmid from induced culture showed supercoiled substrate plasmid, bacterial backbone, seamless vector and open circular seamless vector. In the uninduced culture (absence of recombination), the restriction digestion of the isolated plasmid resulted in linearized substrate plasmid (7.2kb), whereas undigested plasmid showed different conformations of substrate plasmid. L: 1kb ladder; Undig: plasmid isolated from *E. coli* strain LZ41 induced and uninduced culture; Dig: restriction enzyme *EcoRV* digested plasmid isolated from *E. coli* strain LZ41 induced and uninduced culture. The purified 4.1kb seamless vector resulting from *in vivo* recombination was analyzed by agarose gel electrophoresis shown at the right; supercoiled seamless vector: SC-SV; open circular seamless vector: OC-SV

3.1.6. Site-specific seamless vector insertion into human *LINE-1*

3.1.6.1. Seamless vector transgenesis in HT1080 cells

After successfully generating seamless vector via *in vitro* or *in vivo* intra-molecular recombination, the next aim was to achieve site-specific integration of the seamless *attL4X* vector at the genomic *attH4X* site within *LINE-1* elements. For this purpose, a seamless vector (1.8kb) carrying *attL4X* and a puromycin selection cassette as the payload was used for transfection (figure 3.15 a). Human fibrosarcoma HT1080 cells were co-transfected with different amounts of the seamless vector and expression plasmid for either Int-C3 or a recombination-inactive Int variant as shown in table 5. Recombination-inactive Int variant served as a negative control as it carries an alanine substitution

for the scissile bond-forming tyrosine at position 342 (1). Thus, inactive Int cannot catalyse DNA strand cleavage and exchange resulting in inhibition of the integrase mediated recombination reaction. Puromycin selection was initiated 48hrs after transfection and extracted genomic DNA was analysed for integration events via PCR/sequence analysis (figure 3.15 b).

Figure 3.15. Targeting of seamless vector to *LINE-1* elements in HT1080 cells. (a) A schematic illustration showing the targeting of genomic *attH4X* site with seamless vector *attL4X-PGKssPuro* by Int-C3, leading to the integration of the PGK-Puro cassette. Positions of primers used to identify successful integration at *attH4X* sites in *LINE-1* elements are indicated. (b) Experimental strategy demonstrating the λ -Int-mediated targeting of seamless vectors for various cell lines. This includes co-transfection of seamless vectors and Int-expressing plasmids, followed by selection, expansion, and characterization of the targeted clones using PCR, sequencing and functional studies [reprinted with permission (181)].

Table 5. Summary of transfection in HT1080 cells with *attL4X-PGKssPuro* and Int expression plasmids.

Plate#	<i>attL4X-PGKssPuro</i>	Int expression plasmids
1	50ng	500ng <i>pCMVssInt-C3CNLS</i>
2	50ng	500ng <i>pCMVssIna</i>
3	100ng	1 μ g <i>pCMVssInt-C3CNLS</i>
4	100ng	1 μ g <i>pCMVssIna</i>
5	200ng	2 μ g <i>pCMVssInt-C3CNLS</i>
6	200ng	2 μ g <i>pCMVssIna</i>

HT1080 bulk cultures containing cells from ~90-150 initial colonies were obtained from each amount of transfected seamless vector with Int C3 and Inactive-Int. The bulk cultures were expanded and screened by PCR for site-specific insertion into the human genome by using four established diagnostic primer pairs (1) (genomic PCRs were performed by Dr Harshyaa Makhija). These primer pairs were designed to bind the consensus ORF1 sequence of *LINE-1* elements. The forward primers (*cs_attH4X_F1* & *cs_attH4X_F2*) bind ~180-200bp upstream whereas reverse primers (*cs_attH4X_R1* & *cs_attH4X_R2*) bind ~100-180bp downstream of the 18bp *attH4X* sequence to

facilitate PCR screening of specific integration events at *attH4X*. PCR products were detected in all transfected bulk cultures but were absent from control reactions (figure 3.16). The sequence analysis of PCR products confirmed targeted seamless vector insertion into human *LINE-1* and their respective coordinates are shown in table 6. It was observed that different genomic loci were identified for the left and right junction with each bulk culture. This indicated that multiple *LINE-1* elements had been targeted in each bulk culture. However, only specific genomic loci were identified through PCR analysis, most likely due to biased PCR amplification as the DNA templates used for this PCR was obtained from bulk cultures.

Figure 3.16. PCR analysis revealing the integration of seamless vector *attL4X-PGKssPuro* at *attH4X* site in *LINE-1* elements. PCR was performed with primers *cs_attH4X_F1/F2* and Puro Rev 24 (left junction) and *cs_attH4X_R1/R2* and Puro Fwd 127 (right junction), using genomic DNA from three independent transfections with 50ng, 100ng, and 200ng of *attL4X-PGKssPuro* seamless vector. PCR amplified products of the expected size (~1000bp; for the left junction) were detected in bulk C3_50, C3_100 and C3_200 (left panels) and (~950bp; for the right junction) in bulk C3_50, C3_100 and C3_200 (right panels). W, no DNA template control; HT, control with genomic DNA from HT1080 parental cells. Ina_100/Ina_200: bulk genomic DNA from puromycin resistant clones obtained through co-transfection of 100ng and 200ng *attL4X-PGKssPuro* seamless vector and 1 μ g and 2 μ g *pCMVssIna*, respectively. C3_50, C3_100, C3_200: bulk genomic DNA from puromycin resistant HT1080 clones obtained through co-transfection of 50ng, 100ng and 200ng of *attL4X-PGKssPuro* seamless vector with 500ng, 1 μ g, and 2 μ g *pCMVssInt-C3CNLS*, respectively; M1, 1kb DNA ladder [reprinted with permission (181)].

Table 6. List of genomic coordinates and locations of targeted *attH4X* in HT1080 bulk cultures identified using different primer pairs.

Samples	Genomic coordinates	Genomic locations	<i>LINE-1</i> elements
Primers: cs_attH4X_R1 + Puro Fwd 127			
C3_50	chr20:14027192-14027288	Intron of MACROD2 gene	L1PA10
C3_100	chr18:3698874-3698964	Intron of DLGAP1 gene	L1PA7
C3_200	chr20:14027192-14027288	Intron of MACROD2 gene	L1PA10
Primers: cs_attH4X_R2 + Puro Fwd 127			
C3_50	chr3:115752168-115752305	Intergenic	L1PA7
C3_100	chr3:115752168-115752305	Intergenic	L1PA7
C3_200	chr14:25137990-25138154	Intergenic	L1P3
Primers: cs_attH4X_F1 + Puro Rev 24			
C3_50	chr7:36257488-36257573	Intron of EEPD1 gene	L1PA7
C3_100	chr10:44267991-44268090	Intergenic	L1P2
C3_200	chr7:36257468-36257573	Intron of EEPD1 gene	L1PA7
Primers: cs_attH4X_F2 + Puro Rev 24			
C3_50	Insufficient sequencing data		
C3_100	chr3:17450647-17450731	Intron of TBC1D5 gene	L1P3
C3_200	chr15:38572194-38572340	Intergenic	L1PA7

In addition to screening bulk cultures, 92 clones obtained from co-transfection of 50ng seamless vector with 500ng Int-C3 and two clones obtained from co-transfected 50ng seamless vector with 500ng Inactive Int were PCR-screened for seamless vector transgenesis using three established primer pairs (1) (figure 3.17 a). These primers bind to the consensus ORF1 sequence of *LINE-1* elements to facilitate PCR screening of specific integration events at *attH4X*. The PCR screening for the left junction of seamless vector integration with consensus primer cs_attH4X_F1 is shown in figure 3.17 b (contributed by Dr Harshyaa Makhija) and for the right junction, with consensus primers, cs_attH4X_R1 and cs_attH4X_R2 are shown in figure 3.17 c, d. From the 92 clones, almost 50% clones (45 clones) were confirmed by PCR/sequencing analysis that showed specific amplification of transgene/genome recombination products. No specific PCR products could be detected in two clones from Inactive Int (negative control, 3rd and 4th well of figure 3.17 b, c, d), indicating rare selectable illegitimate recombination events led to puromycin resistance. The sequence analysis of the obtained PCR products identified the loci targeted by 45 clones in human *LINE-1* (table 7). It was observed that some clones, such as #27, #41, #45, #46, #80, most likely harboured single copy of seamless vector integration as evident from the same genomic locus identified from PCR analysis of all junctions. However, the remaining clones might carry multiple copies of seamless vector integration as different genomic loci were identified from the left and right junction PCR analysis.

(a)

(b)

(c)

Figure 3.17. Screening for *attH4X/attL4X* recombination events in HT1080 clones. (a) PCR analysis of *attH4X* x *attL4X* recombination events using the genomic DNA from puromycin resistant clones obtained from targeting of seamless vector *attL4X-PGKssPuro* in HT1080 cells. Position of primers *cs_attH4X_R1*, *cs_attH4X_R2*, *cs_attH4X_F1*, Puro Fwd 127 and Puro Rev 24 to identify recombinant junctions after successful integration at *attH4X* in *LINE-1* elements are indicated. (b) PCR with *cs_attH4X_F1* and Puro Rev 24 (~1100bp) detected for 13 clones. (c) PCR screening with *cs_attH4X_R1* and Puro Fwd 127 (~950bp) resulted in 32 positive clones. (d) PCR-amplifications with *cs_attH4X_R2* and Puro Fwd 127 (~1000bp) resulted in 33 positive clones. W, no DNA template control; HT, genomic DNA from HT1080 parental cells; IN, genomic DNA from two puromycin resistant colonies obtained through co-transfection of 50ng of *attL4X-PGKssPuro* and 500ng of *pCMVssIna*; 1-92, genomic DNA from puromycin resistant clones obtained through co-transfection of 50ng of *attL4X-PGKssPuro* and 500ng of *pCMVssInt-C3CNLS*; M1, 1kb DNA ladder [reprinted with permission (181)].

Table 7. List of genomic coordinates and locations of targeted *attH4X* in HT1080 clones identified using different primer pairs as shown in figure 3.17 [reprinted with permission (181)].

Clone #	Genomic coordinates	Genomic locations	<i>LINE-1</i> elements
Primers: cs_attH4X_R1 + Puro Fwd 127			
10	Insufficient sequencing data		
12	chr7:94821074-94821150	Intergenic	L1PA7
15	chr17:35525040-35525120	Intron of SLFN12L gene	L1PA7
17	chr7:54854768-54854843	Intergenic	L1PA7
18	Insufficient sequencing data		
19	Insufficient sequencing data		
21	chr3:66311241-66311321	Intron of SLC25A26 gene	L1PA8
22	chr7:143585724-143585804	Intergenic	L1P3
24	Insufficient sequencing data		
26	chr4:60158615-60158695	Intergenic	L1PA7
27	chr1:230475395-230475453	Intergenic	L1PA7
28	chr4:60158619-60158695	Intergenic	L1PA7
35	chr2:58169727-58169815	Intron of FANCL gene	L1P2
39	chr17:35525032-35525120	Intron of SLFN12L gene	L1PA7
41	chr1:113574938-113575026	Intron of MAGI3	L1PA8
42	chr17:35525040-35525120	Intron of SLFN12L gene	L1PA7
43	chr2:224486485-224486529	Intron of CUL3 gene	L1P3
44	chr3:66311241-66311285	Intron of SLC25A26 gene	L1PA8
45	chr3:156067904-156067987	Intergenic	L1PA7
46	chr17:35302490-35302566	Intergenic	L1PA7
47	chr2:224486454-224486529	Intron of CUL3 gene	L1P3
48	chr17:35302490-35302566	Intergenic	L1PA7
50	chr10:47432585-47432662	Intergenic	L1PA6
59	chr1:183683917-183683997	Intron of RGL1 gene	L1PA7
71	chr14:54406591-54406666	Intron of CDKN3 gene	L1P2
74	chr17:66156714-66156794	Intron of CEP112 gene	L1PA7
77	chr2:232036264-232036346	Intron of DIS3L2 gene	L1PA7
78	chr7:82491494-82491569	Intergenic	L1P2
79	chr11:10424843-10424917	Intergenic	L1PA7
80	chr3:169414717-169414803	Intron of MECOM gene	L1P3
82	chr3:169414720-169414795	Intron of MECOM gene	L1P3
89	chr2:58169739-58169815	Intron of FANCL gene	L1P2
Primers: cs_attH4X_R2 + Puro Fwd 127			
14	chr3:157803268-157803384	Intergenic	L1PA7
15	Insufficient sequencing data		
17	chr7:54854762-54854913	Intergenic	L1PA7
25	chr3:817798-817909	Intron of LINC01266 gene	L1PA7
26	chr7:116744329-116744459	Intron of MET gene	L1PA7
27	chr1:230475373-230475514	Intergenic	L1PA7
39	chrX:44254389-44254542	Intron of EFHC2 gene	L1PA7

40	chr3:25361222-25361370	Intron of RARB gene	L1PA7
41	chr1:113574929-113575086	Intron of MAGI3 gene	L1PA8
42	chr2:6903558-6903715	Intergenic	L1PA7
44	chr3:125997822-125997973	Intergenic	L1P2
45	chr3:156067904-156068021	Intergenic	L1PA7
46	chr17:35302459-35302575	Intergenic	L1PA7
47	chr17:35302465-35302575	Intergenic	L1PA7
48	chr17:35302459-35302580	Intergenic	L1PA7
50	chr2:150193440-150193595	Intron of AC016682.1	L1PA7
52	chr6:23431503-23431654	Intergenic	L1PA8
54	chr12:101061624-101061772	Intron of ANO4 gene	L1PA7
59	chr1:183683908-183684064	Intron of RGL1 gene	L1PA7
65	Insufficient sequencing data		
66	Insufficient sequencing data		
67	chr2:210446661-210446782	Intron of LANCL1-AS1 gene	L1PA7
68	chrX:73893782-73893936	Intergenic	L1PA7
71	chr1:183683908-183684062	Intron of RGL1 gene	L1PA7
74	chr10:31567543-31567656	Intergenic	L1PA7
75	chr10:56762675-56762703	Intergenic	L1M4
76	chr3:157803265-157803418	Intergenic	L1PA7
77	chr2:232036264-232036422	Intron of DIS3L2 gene	L1PA7
78	chr7:82491485-82491620	Intergenic	L1P2
80	chr3:169414652-169414804	Intron of MECOM gene	L1P3
82	chr3:169414655-169414804	Intron of MECOM gene	L1P3
83	chr3:157803265-157803416	Intergenic	L1PA7
87	chr12:88438242-88438391	Intergenic	L1PA7
Primers: cs_attH4X_F1 + Puro Rev 24			
27	chr1:230475281-230475380	Intergenic	L1PA7
39	chrX:44254535-44254610	Intron of EFHC2 gene	L1PA7
40	chr2:39812312-39812427	Intergenic	L1PA7
41	chr1:113574820-113574937	Intron of MAGI3 gene	L1PA8
42	chr2:6903466-6903566	Intergenic	L1PA7
44	chr3:125997966-125998072	Intergenic	L1P2
45	chr3:156067809-156067893	Intergenic	L1PA7
46	chr17:35302567-35302667	Intergenic	L1PA7
48	chr17:35302567-35302679	Intergenic	L1PA7
65	chr3:1033771-1033879	Intergenic	L1PA8
71	chr6:80613523-80613618	Intergenic	L1PA6
77	chr7:36257468-36257573	Intron of EEPD1 gene	L1PA7
80	chr3:169414796-169414904	Intron of MECOM gene	L1P3

In the current study, *attL4X* on the seamless vector was employed as recombination partner instead of *attP4X* for λ Int-mediated recombination at genomic *attH4X* site. Thus, it was interesting to determine different *LINE-1* loci of seamless vector transgenesis across different targeting attempts. For this, a novel PCR based strategy as shown in figure 3.18 a was designed to identify individual targeted *LINE-1* loci (PCR strategy provided by Dr Farid John Ghadessy). The strategy was used on the genomic DNA from bulk HT1080 cell cultures obtained from co-transfection of 100ng of the 1.8kb seamless vector with Int-C3 (C3_100) / Inactive-Int (INA_100). Upon PCR analysis, an expected product was obtained only with C3_100 (figure 3.18 b, contributed by Dr Harshyaa Makhija). Deep paired-end sequencing of the PCR product resulted in 208,936 reads, which contained the expected junction sequence between vector and *attH4X*. This corresponded to 26,380 unique sequences, which were mapped to the human genome (GEO accession: GSE99883). The top hit for each read was tabulated and merged where they overlapped. This yielded 434 different *LINE-1* loci where genomic integration had occurred in bulk cells (Appendix A1, contributed by Dr Shawn Hoon). Almost 3/4th of these loci was counted only thrice or less, which could indicate that these loci were less frequently targeted.

Figure 3.18. Seamless vector genomic insertion site mapping via deep sequencing. (a) An illustration depicting the PCR strategy for mapping insertion sites of seamless vector *attL4X-PGKssPuro* in human HT1080 cells. This includes generation of targeted ssDNA using a biotinylated primer complementary to the inserted vector sequence (step 1), capture of the ssDNA with streptavidin beads, ligation of adenylated oligonucleotide (adaptor) and PCR amplification and sequencing analysis (steps 2-6). Positions of the relevant primers used to map *attH4X* (right junction) are indicated in

step 5. (b) PCR analysis for mapping insertion sites from HT1080 bulk cell culture. PCR with the primers bpa1 and cs_attH4X_R1 (see step 5 in Figure 3.18 a), using bulk genomic DNA from co-transfections of 100ng of *attL4X-PGKssPuro* and *pCMVssInt-C3CNLS* resulted in a specific band (highlighted), that was not observed with bulk genomic DNA from control co-transfections of 100ng of *attL4X-PGKssPuro* and *pCMVssIna*. M1, 1kb ladder; HT, genomic DNA from parental HT1080 cells; INA_100: bulk genomic DNA from co-transfection of 100ng of *attL4X-PGKssPuro* and 1µg Inactive Int; C3_100: bulk genomic DNA from co-transfection of 100ng of *attL4X-PGKssPuro* and 1µg Int-C3. W: no DNA template control [reprinted with permission (181)].

3.1.6.2. *In vivo* sequential intra- and inter-molecular recombination leading to transgenesis of seamless vectors

To explore the different possible recombination scenarios for co-transfection of substrate plasmid with Int-C3, a substrate plasmid carrying direct repeats *attP4X* and *attH4X* and a puromycin expression cassette as the payload was used for transfection in HT1080 cells. There are two possible recombination scenarios that might occur with this approach (figure 3.19). Firstly, the substrate plasmid carrying *attP4X* can recombine with genomic *attH4X* in the presence of Int-C3, leading to the insertion of the entire substrate plasmid into human *LINE-1* flanked by *attL4X* and *attR4X* (left pathway). Secondly, intramolecular recombination between *attH4X* and *attP4X* on the substrate plasmid might occur first in the presence of Int-C3, resulting in a seamless vector carrying *attL4X*. The seamless vector can subsequently get integrated into human *LINE-1* by a second recombination reaction with Int-C3 between *attL4X* and genomic *attH4X* (right pathway).

Figure 3.19. A schematic illustration showing *in vivo* intramolecular recombination between *attH4X* and *attP4X* in substrate plasmid. Left panel demonstrates the intermolecular recombination between *attP4X* and genomic *attH4X* resulting integration of the substrate plasmid in human *LINE-1* element flanked by *attL4X* and *attR4X*. Right panel demonstrates the intramolecular recombination between *attH4X* and *attP4X* in substrate plasmid resulting in the formation of the seamless vector carrying *attL4X* which recombines with genomic *attH4X* leading to insertion of a seamless vector in the human *LINE-1* element. Positions of relevant primers Puro rev24, PGK Rev, Puro fwd 127, Puro fwd 128, PR21, 196, cs_attH4X_F1/F2 and cs_attH4X_R1/R2 used to check the integration at the *attH4X* site in the *LINE-1* element are indicated.

In order to test these pathways, HT1080 cells were co-transfected with substrate plasmid *pattP4X-PGKssPuro-attH4X* (5kb) and Int expression plasmids Int-C3 or inactive Int. Puromycin selection was initiated 48hrs after transfection, and extraction of genomic DNA was carried out for PCR/sequence analysis. Thirty clones obtained from co-transfection of 100ng substrate plasmid with 1µg Int-C3 and six clones obtained from co-transfection of 100ng substrate plasmid with 1µg Inactive Int were expanded. Based on the specificity of PCR primers for the two expected right junction outcomes (Fig. 3.17), it was found that 13/30 clones (43%) carried an integration of the seamless vector into human *LINE*, as a result of intramolecular recombination between *attP4X* and *attH4X* before the intermolecular recombination (figure 3.20 a, b). Furthermore, 6 clones were identified that resulted from substrate plasmid integration into human *LINE-1* generated by intermolecular *attP4X* (substrate plasmid) and a genomic *attH4X* site recombination events leading to the *attR4X* junction sequence (figure 3.20 c, d). However, no specific PCR products could be detected in rest of 12 clones and also in 6 clones obtained from Inactive Int (negative control, A-G lanes of figure 3.20 a, b, c, d, e, f). This might happen due to random insertion of the substrate plasmid/seamless vector in the HT1080 genome that led to puromycin resistance. The left junction of both seamless vector and substrate plasmid after recombining with *LINE-1* is identical and was confirmed in 11 clones by PCR/sequencing (figure 3.20 e, f). The sequence analysis identified the targeted *LINE-1* loci in all clones, as listed in table 8. Interestingly, it was found that clone #7 harboured both substrate plasmid and seamless vector integration at two different genomic loci into human *LINE-1* (table 8).

(a) *cs_attH4X_R1* + Puro Fwd 128

(b) *cs_attH4X_R2* + Puro Fwd 128

Figure 3.20. Screening for *attH4X/attL4X* and *attH4X/attP4X* recombination events in HT1080 cells. PCR screening of *in vivo* intramolecular recombination between *attP4X* and *attH4X* using the genomic DNA from puromycin resistant clones obtained through co-transfection of 100ng of *attP4X-PGKssPuro-attH4X* and 1μg of *pCMVssInt-C3CNLS*. a) A total of 13/30 puromycin resistant clones were identified positive for *LINE-1* targeting by seamless vector via *attL4X* and genomic *attH4X* recombination in an initial screen involving semi-nested PCR with primers cs_attH4X_R1, cs_attH4X_R2, Puro Fwd 128 using the genomic DNA template from primary PCR (with primers Puro Fwd 127) as indicated. Recombinant *attH4X* junction (generated from *attL4X* x *attH4X*) screened by PCR with cs_attH4X_R1 and Puro Fwd 128 (~760bp) resulted in 8 positive clones. b) PCR-amplifications with cs_attH4X_R2 and Puro Fwd 128 resulted in 10 positive clones. c) A total of 6/30 puromycin resistant clones were identified positive for *LINE-1* targeting by substrate plasmid via *attP4X* and genomic *attH4X* intermolecular recombination in an initial screen involving semi-nested PCR with primers cs_attH4X_R1, cs_attH4X_R2, PR 21 using the genomic DNA template from primary PCR (with primers 196) as indicated. The *attR4X* junction (generated from *attP4X* x *attH4X*) screened by PCR with cs_attH4X_R1 and PR 21 (~380bp) resulted in 6 positive clones. d) PCR-amplifications with cs_attH4X_R2 and PR 21 resulted in 6 positive clones. e) Recombinant *attL4X* junction (generated from *attL4X* x *attH4X* or *attP4X* x *attH4X*) recombination was screened by semi-nested PCR with primers cs_attH4X_F1/F2 and PGK Rev using the genomic DNA template from primary PCR (with primers Puro Rev 24) as indicated. PCR-amplifications with cs_attH4X_F1 and Puro

Rev 24 (~930bp) were detected for 8 clones. f) PCR screening with *cs_attH4X_F2* and PGK Rev resulted in 9 positive clones W, no DNA template control; HT, genomic DNA from parental HT1080 cells; A-G, genomic DNA from puromycin resistant clones obtained through co-transfection of 100ng of *attP4X-PGKssPuro-attH4X* and 1µg of *pCMVssIna*; 1-30, genomic DNA from puromycin resistant clones obtained through co-transfection of 100ng of *attP4X-PGKssPuro-attH4X* and 1µg of *pCMVssInt-C3CNLS*; L, 1kb DNA ladder.

Table 8. List of genomic coordinates and locations obtained upon targeting of either substrate plasmid or seamless vector in HT1080.

Clone #	Genomic coordinates	Genomic locations	<i>LINE-1</i> elements
Primers: <i>cs_attH4X_R1</i> + Puro Fwd 128			
2	chr4:71310048-71310113	Intron of SLC4A4 gene	L1PA8
5	chr5:167959432-167959512	Intron of TENM2 gene	L1PA7
7	chr2:160182356-160182423	Intron of ITGB6 gene	L1PA7
9	chr20:35162942-35162997	Intron of mRNA JD056712	L1PA7
18	chr5:167959432-167959489	Intron of TENM2 gene	L1PA7
24	chr12:88668820-88668884	Intergenic	L1PA7
25	chr5:167959432-167959496	Intron of TENM2 gene	L1PA7
30	chr1:113574920-113574997	Intron of MAGI3 gene	L1PA8
Primers: <i>cs_attH4X_R2</i> + Puro Fwd 128			
2	chr12:88668759-88668884	Intergenic	L1PA7
3	chr12:88668759-88668884	Intergenic	L1PA7
4	chr12:88668759-88668884	Intergenic	L1PA7
5	chr12:88668759-88668884	Intergenic	L1PA7
7	chr2:160182296-160182425	Intron of ITGB6	L1PA7
9	chr10:52373960-52374085	Intron of mRNA AK130042	L1P3
15	chr12:88668760-88668884	Intergenic	L1PA7
20	chr2:181090524-181090651	Intron of AC068196.1	L1PA7
24	chr12:88668759-88668884	Intergenic	L1PA7
28	chr2:150193481-150193604	Intron of AC016682.1	L1PA7
Primers: <i>cs_attH4X_R1</i> + PR21			
7	chr4:86192505-86192569	Intron of MAPK10 gene	L1PA7
8	chr12:105012219-105012281	Intergenic	L1PA7
10	chr11:40470758-40470814	Intron of LRRC4C gene	L1PA7
22	chr4:86192505-86192569	Intron of MAPK10 gene	L1PA7
23	chr10:27588682-27588743	Intergenic	L1PA7
26	chr10:27588680-27588743	Intergenic	L1PA7
Primers: <i>cs_attH4X_R2</i> + PR21			
7	chr4:86192505-86192646	Intron of MAPK10	L1PA7
8	chr12:105012219-105012361	Intergenic	L1PA7
10	chr10:27588588-27588743	Intergenic	L1PA7
22	chr4:86192505-86192659	Intron of MAPK10	L1PA7
23	chr10:27588601-27588743	Intergenic	L1PA7
26	chr10:27588601-27588743	Intergenic	L1PA7
Primers: <i>cs_attH4X_F1</i> + PGK Rev			
7	chr12:1941144-1941257	Exon of mRNA LF207361	L1P2
8	chr12:105012135-105012236	Intergenic	L1PA7
10	chr10:27588726-27588839	Intergenic	L1PA7
18	chr6:121704332-121704440	Intergenic	L1PA6

22	chr4:86192416-86192522	Intron of MAPK10	L1PA7
23	chr10:27588726-27588840	Intergenic	L1PA7
26	chr10:27588726-27588833	Intergenic	L1PA7
28	chr2:150193587-150193693	Intron of AC016682.1	L1PA7
Primers: cs_attH4X_F2 + PGK Rev			
7	chr12:1941118-1941257	Exon of mRNA LF207361	L1P2
10	chr10:27588726-27588874	Intergenic	L1PA7
20	chr2:181090634-181090783	Intron of AC068196.1	L1PA7
21	chr7:17808367-17808528	Intron of SNX13	L1PA7
22	chr4:86192372-86192522	Intron of MAPK10	L1PA7
23	chr10:27588726-27588874	Intergenic	L1PA7
25	chr14: 81542069-81542178	Intergenic	L1PA5
26	chr10:27588726-27588885	Intergenic	L1PA7
28	chr2:150193587-150193729	Intron of AC016682.1	L1PA7

3.1.6.3. Seamless vector transgenesis in human embryonic stem cells (hESCs)

To further validate the seamless vector transgenesis tool in a different cell type, targeting of the seamless vector was attempted in human embryonic stem cells (hESCs). hESCs are an attractive potential cell candidate for cell/gene therapy and regenerative medicine due to its self-renewing capability and ability to be differentiated into derivatives of all three germ layers (275). Thus, engineering the genome of hESCs with λ Int-mediated transgenesis tool can provide a proof of concept for its potential application in the gene therapy. For this purpose, hESCs were targeted using the 1.8kb seamless vector carrying *attL4X* and a puromycin expression cassette as payload. hESCs were cotransfected with seamless vector DNA (100ng) and 1 μ g of an expression plasmid for either Int-C3 or a recombination-inactive Int variant as shown in table 9 (transfection was performed by Dr Harshyaa Makhija). After 2 weeks of puromycin selection, 32 hES clones were expanded from seamless vector co-transfection with 1 μ g Int-C3. No viable hES clones were obtained from seamless vector co-transfection with 1 μ g Inactive-Int. PCR analysis of 32 clones was carried out using four established primer pairs (figure 3.21 a) that were designed previously to bind the consensus ORF1 sequence of *LINE-1* elements. The PCR screening for the left junction of seamless vector integration with consensus primers cs_attH4X_F1 and cs_attH4X_F2 is shown in figure 3.21 b, c and for the right junction, with consensus primers, cs_attH4X_R1 and cs_attH4X_R2 is shown in figure 3.21 d, e. Out of 32 clones, 25 clones were identified by PCR and sequencing for seamless vector integration at *LINE-1* in hESCs as shown in table 10.

Table 9. Summary of transfection in hESCs with *attL4X-PGKssPuro* and Int expression plasmids.

Plate#	<i>attL4X-PGKssPuro</i>	Int expression plasmids
1	100ng	1 μ g <i>pEFssInt-C3CNLS</i>
2	100ng	1 μ g <i>pEFssInt-INA</i>

19	chrX:44254483-44254552	intron of EFHC2 gene	L1PA7
27	chr20:13053749-13053819	intron of SPTLC3	L1P3
31	chr7:54854750-54854843	Intergenic	L1PA7
Primers: cs_attH4X_R2 + Puro Fwd 509			
1	chr12:72781550-72781631	Intergenic	L1P3
2	chr12:101061615-101061782	intron of ANO4 gene	L1PA7
3	chr12:101061615-101061765	intron of ANO4 gene	L1PA7
4	chr10:31567534-31567630	Intergenic	L1PA7
5	chr12:101061615-101061782	intron of ANO4 gene	L1PA7
7	chr3:157803256-157803426	Intergenic	L1PA7
8	chr11:96488217-96488387	Intron of JRKL-AS1 gene	L1PA6
9	chr12:101061615-101061782	intron of ANO4 gene	L1PA7
10	chr12:101061615-101061782	intron of ANO4 gene	L1PA7
11	chr3:157803256-157803426	Intergenic	L1PA7
12	chr3:157803256-157803426	Intergenic	L1PA7
13	chr2:73509950-73510120	Intron of ALMS1 gene	L1PA7
14	chr17:35302478-35302584	intergenic	L1PA7
15	chr3:157803256-157803426	Intergenic	L1PA7
16	chr3:157803256-157803426	Intergenic	L1PA7
19	chr7:54854750-54854920	Intergenic	L1PA7
20	chr3:157803256-157803426	Intergenic	L1PA7
27	chr3:157803256-157803413	Intergenic	L1PA7
29	chr12:101061615-101061782	intron of ANO4 gene	L1PA7
30	chr3:157803256-157803426	Intergenic	L1PA7
32	chr12:44898931-44899037	intron of NELL2	L1PA7
Primers: cs_attH4X_F1 + PGK Rev			
3	chr3:156067764-156067893	Intergenic	L1PA7
Primers: cs_attH4X_F2 + PGK Rev			
7	chr2:159357384-159357465	Intron of BAZ2B gene	L1PA7
15	chr2:159357384-159357515	Intron of BAZ2B gene	L1PA7
17	chr2:159357384-159357490	Intron of BAZ2B gene	L1PA7
29	chr2:159357384-159357485	Intron of BAZ2B gene	L1PA7
32	chr2:159357384-159357505	Intron of BAZ2B gene	L1PA7

3.1.7. Application of seamless vector in cancer gene therapy

Seamless vectors as episomal DNA vectors provide several advantages over conventional plasmid DNA vectors. It has been found that seamless vector-encoded transgene cassette expression is about three to ten folds higher than the expression achieved with parental plasmids (192). Thus, seamless vector transgenesis has been under consideration as a promising approach to circumvent the safety issues associated with conventional plasmids and to ensure high expression of the transgene for therapeutic applications (276). Seamless vectors have been studied for years in pre-clinical research for the treatment of a wide range of diseases like hemophilia B, AAT deficiency, myocardial

infarction, cystic fibrosis, and peripheral arterial diseases (194,277,278). They have also been used to produce a protective DNA vaccine against listeriosis and antitumoral molecule against cancer cells (279,280). Seamless vectors showed enhanced therapeutic effects in clinical cases when compared to conventional plasmids, and, hence, can be a safer alternative to plasmid vector for next-generation biopharmaceuticals, cell therapy, DNA vaccination and gene therapy (276).

One such application for seamless vectors generated by λ -Int that has been explored in the study is the cancer immunotherapy, which is an emerging field for potent and powerful treatment of cancer using genetically engineered tumour-reactive T cells. This involves insertion and expression of synthetic chimeric antigen receptor (CAR) on T cells against specific antigens of tumours (281). Chimeric antigen receptors (CAR) are fusion proteins that comprise of an antigen-binding domain consisting of variable light chains of a monoclonal antibody, a single-chain variable fragment (scFv) of variable heavy (VH) and an intracellular signalling module (282) (figure 3.22 a). Different transmembrane domains and spacers are used to link the intracellular and extracellular portions that anchor the receptor on the T cell surface. The first-generation CARs contained only one signalling module, the CD3 ζ chain (Signal 1), which further evolved into second- and third-generation CARs having two or three (respectively) costimulatory domains, such as ICOS, 4-1BB, CD28, OX40, to deliver Signal 2 (282). This signal is crucial for better proliferation, higher cytokine production, improved survival and optimal T cell stimulation to mediate anti-tumoral activity (figure 3.20 a) (283). The CAR-T cells can recognize antigens without MHC presentation. It mediates CD19+ malignant B-cells killing by enabling T cells to bind to the CD19 surface antigen using the scFv recognition domain. Upon interaction, CAR-T cells exhibit a non-classical immune synapse (IS) and finally facilitate the tumour cells killing by perforin/granzyme cytotoxicity, activating the Fas and FasL ligand pathway, along with the cytokine release to sensitize the tumour stroma (284) (figure 3.22 b).

Figure 3.22. Structure and function of Chimeric antigen receptors (CAR). (a) Structure of first, second and third-generation CAR having different signaling molecules (285) b) Tumour killing mechanism of CAR T cells [reprinted with permission (284)].

Current technology for expression of CAR transgene on T cells involves a) transient expression by electroporating CAR-encoding mRNA in T cells (286,287) (b) CAR transgene transfer by using viral or transposon-based vectors (288,289). Both technologies suffer from severe limitations. The transient expression of CAR by mRNA lasts for only 5-7 days due to the degradation of mRNA during T cell proliferation. Hence, the CAR-T cells are short-lived which leads to the requirement of multiple rounds of CAR-T cells infusions in patients to achieve any therapeutic actions (290,291). In contrast, permanent expression by stable integration of CAR transgene by viral or transposon-based gene transfer methods had resulted in insertional mutagenesis, remobilization of transposons in T cells (288,289) and other unwanted side effects such as cytokine release syndrome (CRS) (292). In addition, the process to generate the viral vectors encoding CAR is time-consuming and expensive (29). Even though CAR-T cells have successfully provided therapeutic benefits, yet the high cost of generation of autologous CAR-T cell makes its usage economically impractical (293). One alternative solution is to use allogenic T cells derived from induced pluripotent stem cells (iPSCs) or human embryonic stem cells (hESCs). Development of such an approach can lead to the formation of a T cell source bank to mitigate its high demand and also reduce the cost of manufacture (293,294). In this respect, λ -Integrase was utilized here to generate a seamless vector encoding anti-CD19 CAR followed by *ex vivo* site-specific insertion and expression of the anti CD19 CAR transgene in hESCs as a proof of concept for exploring the λ -Integrase transgenesis tool for therapeutic protein expression.

3.1.8. Anti-CD19 CAR seamless vector insertion and sustained expression in hESCs

The aim was to deliver a transgene construct that expressed a second-generation anti-CD19 CAR (provided by Prof. Dario Campana) as payload to hESCs (295). For this purpose, a seamless vector,

attL4X-PGKssPuro-EF-CAR, was generated through *in vitro* recombination from *pattP4X-PGKssPuro-EF-CAR-attH4X* (figure 3.23 a). hESCs were co-transfected with 5.5kb seamless vector *attL4X-PGKssPuro-EF-CAR* (100ng) and plasmids expressing Int-C3 (1 μ g) that resulted in puromycin resistant clones which were screened for seamless vector insertion and expression. In parallel, co-transfection between 5.5kb seamless vector (100ng) and plasmids expressing inactive Int (1 μ g) was performed as a control. Seventy-two hES clones obtained from co-transfection of 5.5kb seamless vector and Int-C3 expressing plasmid were expanded. Simultaneously, 6 hES clones obtained from co-transfection of the seamless vector with inactive Int were also allowed to expand. PCR and sequencing analysis using two established primer pairs, that were designed previously to bind the consensus ORF1 sequence of *LINE-1* elements, showed 2 clones (out of 72 clones) have seamless vector insertion in hESCs genome, while no insertion was observed in control (figure 3.23 b, c). Next, anti-CD19 CAR expression was analysed in two positive hES clones through flow cytometry by staining the cells with anti-human CD3 ζ primary antibody (contributed by Dr Harshyaa Makhija). Surprisingly, anti-CD19 CAR expression was not detected in flow cytometry of these clones even though PCR analysis for seamless vector amplification showed that the complete integration of *PGKssPuro-EF-CAR* cassette into *LINE-1* elements (data not shown). This could be because the hESCs genome is highly heterochromatinized (296,297). There is a possibility that seamless vector was integrated into the hESC heterochromatinized genome and thus due to epigenetic silencing of that region, the transgene cassette was not expressed.

Figure 3.23. Targeting of a two-reporter seamless vector *attL4X-PGKssPuro-EF-CAR* in hESCs. (a) A schematic drawing showing *in vitro* generation and *ex vivo* targeting of a two-reporter seamless vector (*attL4X-PGKssPuro-EF-CAR*) at genomic *attH4X*. Positions of relevant screening primers Puro rev 24, CAR Fwd, *cs_attH4X_F2/R2* are indicated. The purified 5.5kb seamless vector resulting from *in vitro* recombination was analyzed by agarose gel electrophoresis shown at the left; supercoiled seamless vector: SC-SV; open circular seamless vector: OC-SV. (b) PCR screening of *attL4X* x *attH4X* recombination events using the genomic DNA from puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-PGKssPuro-EF-CAR* and 1μg of *pEFssInt-C3CNLS*. A total of 2/72 clones were identified positive for recombinant *attH4X* junction (generated from *attL4X* x *attH4X*) screened by PCR with *cs_attH4X_R2* and CAR fwd (~1300bp). (c) Recombinant *attL4X* junction of the two positive clones (generated from *attL4X* x *attH4X*) screened by PCR with *cs_attH4X_F2* and Puro Rev 24 (~1175bp). W, no DNA template control; ES genomic DNA from parental hESCs; 1-72, puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-PGKssPuro-EF-CAR* and 1μg of *pEFssInt-C3CNLS*; IN, genomic DNA from puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-PGKssPuro-EF-CAR* and 1μg of *pEFssInt-Ina*, 1kb DNA ladder.

In parallel, another plasmid *pattP4X-PGKss-CAR-EF-Puro-attH4X* was used and seamless vector *attL4X-PGKss-CAR-EF-Puro* (5.5kb) was produced via *in vitro* recombination (figure 3.24 a). In this case, the promoter to express anti-CD19 CAR transgene was switched from EFα to PGK promoter to obtain enhanced and stable transgene expression in undifferentiated hESCs (298). The seamless vector (100ng) was co-transfected with expression plasmids Int-C3/inactive Int (500ng each) in hESCs (transfection was performed by Dr Harshyaa Makhija). After 2 weeks of puromycin selection, 48 hES clones obtained from seamless vector co-transfection with Int-C3 and 10 hES clones from vector co-transfection with Inactive Int were expanded. PCR and sequence analysis confirmed recombinant *attL4X* junction (generated from *attL4X* x *attH4X*) in 12 hES clones that had been targeted in *LINE-1* loci (figure 3.24 b, c; table 11). Out of these 12 hES clones, only two clones, clone #19 and clone#34 showed anti-CD19-CAR expression, as determined through flow cytometry by staining the cells with anti-human CD3ζ primary antibody (contributed by Dr Harshyaa Makhija). The FACS histogram clearly distinguishes the anti-CD19 CAR-expressing population from non-

transfecting cells (figure 3.24 d). The FACS histogram overlay of non-transfected control population onto inactive Int population showed the same peak suggesting no detectable anti-CD19-CAR expression in inactive Int population. FACS histogram of the anti-CD19 CAR-expressing population of clone #19 and clone #34 showed a separate peak upon overlay onto non-transfected control population indicating 87.1% and 73.3% of anti-CD19 CAR-expressing cells respectively (figure 3.24 d). However, clone#34 did not survive in the cell culture condition, thus, clone#19 was selected for further analysis.

Figure 3.24. Anti-CD19 CAR expression from seamless vectors in hESCs. (a) A schematic drawing showing *in vitro* generation and ex vivo targeting of a two-reporter seamless vector (*attL4X-PGKssCAR-EF-Puro*) at genomic *attH4X*. Positions of relevant screening primers PGK Rev 187, attP Rev and cs_attH4X_F1/F2 are indicated. The purified 5.5kb seamless vector resulting from *in vitro* recombination was analysed by agarose gel electrophoresis shown at the left; supercoiled seamless vector: SC-SV; open circular seamless vector: OC-SV. (b) PCR screening of *attH4X* x *attL4X* recombination events using the genomic DNA from puromycin resistant clones obtained from targeting of seamless vector *attL4X-PGKssCAR-EF-Puro* in hESCs. The recombinant *attL4X* junction at *LINE-1* elements was identified in a screen involving semi-nested PCR with primers cs_attH4X_F1, cs_attH4X_F2, attP Rev using the genomic DNA template from primary PCR (with primers PGK Rev 187) as indicated. (b) PCR with cs_attH4X_F1 and attP Rev (~220bp) resulted in 10 positive clones. (c) PCR screening with cs_attH4X_F2 and attP Rev (~280bp) resulted in 2 positive clones. W, no DNA template control; ES, genomic DNA from parental hESCs; I1-I10, puromycin resistant clones obtained through co-

transfection of 100ng of *attL4X-PGKssCAR-EF-Puro* and 500ng of *pEFssInt-INA*; 1-48, puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-PGKssCAR-EF-Puro* and 500ng of *pEFssInt-C3CNLS*; L, 1kb DNA ladder. (d) Flow cytometric analysis of anti-CD19 CAR expression in targeted hESC clones#19 and#34. A single parameter histogram overlay plots representation of stained non-transfected control hESCs onto the anti-CD19 CAR-stained population resulting from co-transfection of anti-CD19 CAR seamless expression vector with Int-C3 expression plasmid; clone#19 (middle panel) and clone#34 (right panel). An overlay obtained with cells from co-transfection of anti-CD19 CAR seamless expression vector with inactive Int expression plasmid and non-transfected control hESCs after staining is shown in the left panel indicate near absence of expression. The overlay of non-transfected control population onto the transfected population differentiate the anti-CD19 CAR-expressing population seen with Int-C3-transfected clones# 19 and #34 [reprinted with permission (181)].

Table 11. Location of *attH4X* sequences in hESC clones targeted with anti-CD19-CAR seamless vector identified using different primer pairs as shown in figure 3.24 a [reprinted with permission (181)].

Clone #	Genomic coordinates	Genomic locations	LINE-1 elements
Primers: cs_attH4X_F1 + attP Rev			
1	chr12:88438383-88438489	Intergenic	L1PA7
2	chr12:88438383-88438495	Intergenic	L1PA7
3	chr12:88438383-88438496	Intergenic	L1PA7
23	chr3:156067805-156067912	Intergenic	L1PA7
34	chr2:153327436-153327543	Intergenic	L1PA7
36	chr3:156067805-156067911	Intergenic	L1PA7
39	chr12:88438384-88438497	Intergenic	L1PA7
46	chr12:88438384-88438489	Intergenic	L1PA7
47	chr12:88438383-88438491	Intergenic	L1PA7
48	chr12:88438384-88438498	Intergenic	L1PA7
Primers: cs_attH4X_F2 + attP Rev			
19	chr2:159357384-159357546	Intron of BAZ2B gene	L1PA7
25	chr2:159357384-159357549	Intron of BAZ2B gene	L1PA7

3.1.8.1. Western blot analysis of anti-CD19 CAR expressing clone #19

The anti-CD19 CAR expression from clone #19 was further confirmed by western blot analysis using a primary antibody against the intracellular CD3 ζ signaling module. As a positive control, a bulk culture of Jurkat T cells obtained from targeting *attL4X-PGKss-CAR-EF-GFP* seamless vector (expressing anti-CD19 CAR) along with plasmid expressing Int C3 was used. Western blot analysis resulted in double bands at ~55 kDa for both clone #19 and Jurkat T cells suggesting post-translational modifications of anti-CD19 CAR protein. Since both clone #19 and Jurkat T cells were being transfected with the same seamless vector that expresses anti-CD19 CAR, western blot showed a similar band pattern at ~55 kDa. An additional band at ~15kDa band was observed for Jurkat T cells that indicated the expression of the native monomer of endogenous CD3 ζ (Figure 3.25) which was absent in hESCs. No signal was observed for non-transfected hESCs and inactive clone.

Figure 3.25. Western blot showing anti-CD19 CAR expression. Western analysis using anti-human CD3 ζ primary antibody and a goat anti-mouse IgG horseradish peroxidase-conjugated secondary antibody shows anti-CD19 CAR expression in clone #19. (Top Panel). β -actin was used as a loading control (bottom panel). Lanes: ES, lysate from non-transfected hESCs; 19, lysate from clone #19 (a puromycin resistant clone obtained from targeting of seamless vector *attL4X-PGKssCAR-EF-Puro* and plasmid expressing Int C3); IN, lysate from an inactive clone (a puromycin resistant clone obtained from targeting of seamless vector *attL4X-PGKssCAR-EF-Puro* and plasmid expressing inactive integrase); JUR, a bulk culture of Jurkat T cells obtained from targeting *attL4X-PGKss-CAR-EF-GFP* seamless vector (expressing anti-CD19 CAR) along with plasmid expressing Int C3 was used as a positive control [reprinted with permission (181)].

3.1.8.2. Southern blot analysis of anti-CD19 CAR expressing clone #19

To analyze the copy number of the integrated transgene cassette, Southern blot analysis was carried out using probes complementary to *CAR* or *Puro* (figure 3.26 a). As shown in figure 3.24 b, c, there is only one band present for clone #19 with both probes complementary to *CAR* and *Puro*, suggesting that the targeted clone#19 carried a single-copy transgene of indicated sizes. However, a single band was also observed for inactive clone (obtained from co-transfection of *attL4X-PGKssCAR-EF-Puro* and plasmid expressing inactive integrase) with both the probes (figure 3.26 b, c) indicating random insertion of a single copy of the seamless vector in the hESC genome. It was noticed that corresponding genomic fragment size obtained by Southern analysis of clone #19 showed a discrepancy with the sequence information obtained from the UCSC browser (table 12). Similar discrepancies were also encountered in a previous study (1) which might be due to two possibilities. First, the genome annotation of the corresponding *LINE-1* elements may not be entirely correct perhaps due to the repetitive nature of these sequences. Second, sequence polymorphism surrounding these sequences could generate different restriction patterns in different genetic backgrounds (299-303).

Figure 3.26. Southern blot analysis of the targeted hESC clone#19. (a) A schematic drawing showing probe complementary to *CAR* and *Puro* on the seamless vector (*attL4X-PGKssCAR-EF-Puro*) targeted at genomic *attH4X*. Positions of relevant restriction sites for *BsrGI*, *NcoI* and *BamHI* are indicated. A southern blot analysis showing the bands indicated by arrows obtained using PCR-derived digoxigenin-labelled probe complementary to *CAR* (b) or *Puro* (c) with genomic DNA purified from targeted hESC clone#19 digested with *BsrGI*, *NcoI* and/or *BamHI*. Lanes: pPCEP (10⁸, 10⁷), copies of linearized vector *pattP4X-PGKssCAR-EF-Puro-attH4X* loaded as positive control; ES, genomic DNA from parental DNA; IN, genomic DNA obtained from inactive clone (a puromycin resistant clone obtained from co-transfection of *attL4X-PGKssCAR-EF-Puro* and plasmid expressing inactive integrase); 19, genomic DNA from targeted hESC clone. hESC clone#19 (carrying a single-copy transgene) is a puromycin resistant clone obtained through co-transfection of *attL4X-PGKssCAR-EF-Puro* and plasmid expressing active Int C3 [reprinted with permission (181)].

Table 12. Summary of Southern blot analysis for targeted clone #19 using probes complementary to *CAR* and *Puro*

Probe	Restriction digest	Expected fragment size (103)	Observed fragment size (103)
<i>CAR</i> gene	<i>BsrGI</i>	>10	~5.5
<i>Puro</i> gene	<i>NcoI</i>	6.1	~9
	<i>BsrGI</i>	>10	~5.3
	<i>BamHI</i>	2.4	~20

3.1.8.3. Long term and stable anti-CD19 CAR expression in clone #19

The anti-CD19 CAR expression of targeted clone#19 was analysed after 6 months and it showed stable transgene expression in 90% of the cells (figure 3.27, right panel) as compared to 87% of the cells in the 1st month of culture (figure 3.27, left panel). This indicated remarkably stable and long term anti-CD19 CAR expression from targeted clone #19. Interestingly, the fraction of low anti-

CD19-CAR expressing cells substantially decreased over 6 months, perhaps indicating that some sort of suppression of negative epigenetic modulation of transgene expression at this locus over time.

Figure 3.27. Long-term and stable anti-CD19 CAR expression from seamless vectors in hESCs. An overlay obtained with clone #19 (after 1st month of culturing) and wild-type hESCs differentiates the anti-CD19 CAR-expressing population as is shown in the left panel. The anti-CD19 CAR expression of clone#19 after 6 months is shown in the right panel [reprinted with permission (181)].

3.1.9. Identification of *LINE-1* as preferred target loci

In order to determine preferred *LINE-1* loci for Int-C3 mediated seamless vector transgenesis, the data from five lists of targeted *attH4X* sequences obtained in HT1080 cells and hESCs (tables 7, 8, 10, 11, Appendix A1) were analyzed. It resulted in the identification of 27 loci, termed as ‘hot-spot’, which were targeted at least twice in four independent cell transfections (table 13). Furthermore, the data were analyzed separately for each cell line (highlighted by different colors in table 13) as the preferential integration sites might vary in different cell types. As evident from table 13, almost half of the identified *attH4X* sequences are in intergenic regions of the human genome, while the rest belongs to introns. Considering that only one-third of the 933 *attH4X*-containing *LINE-1* is located in introns (181), it could be suggested that there is a slight preference of Int-mediated recombination at *attH4x* sequences that belongs to introns. However, one intergenic *attH4X* locus was identified that had been targeted in all four independent targeting events in HT1080 cells and hESCs (highlighted in red, table 13). To further determine the DNA sequence of preferred *attH4X*, a sequence logo was generated on these 27 ‘hot-spot’ *attH4X* sites. The sequence logo (below table 13) showed that among the first three variable nucleotides at the 5’end of 21bp *attH4X* sequence, adenine is slightly preferred at first and third position, while cytosine was preferred at the second position. Interestingly, the cytosine at the second position within *attH4X* appears to play a more important role in genomic recombination reactions with *attL4X*, when compared with the previous study where *attP4X* was used as recombination partner (1).

Table 13. List of 27 “hotspot” *LINE-1*: Genomic coordinates, corresponding *attH4X* sequences, locations and *LINE-1* elements of the “hotspots” compiled from tables 7, 8, 10, 11, Appendix A1. A sequence logo obtained from the 27 “hotspot” sequences is shown at the bottom [adapted with permission (181)].

Genomic coordinates	<i>attH4X</i> sequences	Genomic locations	<i>LINE-1</i> elements	HT1080 clones (Table 7, 8)	Counts (Appendix A1)	hESC clones (Table 10)	CAR targeted hESC clones (Table 11)
chr10:47432585-47432662	TCTCTTTATTTTCATTAAGTTG	Intergenic	L1PA6	1	14116	0	0
chr3:156067904-156068021	ATGCTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	57	1	2
chr7:94821074-94821150	GTGCTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	31	0	0
chr1:230475395-230475453	TCACTTCATTTTCATTAAGTTG	Intergenic	L1PA7	1	10	0	0
chr12:88438242-88438391	CTGCTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	10	0	7
chr7:82491494-82491569	TCACTTTATTTTCATTAAGTTG	Intergenic	L1P2	1	2	0	0
chrX:73893782-73893936	TCACTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	2	0	0
chr3:157803256-157803339	TTGCTTTATTTTCATTAAGTTG	Intergenic	L1PA7	3	0	8	0
chr17:35302478-35302584	ACACTTTATTTTCATTAAGTTG	Intergenic	L1PA7	3	0	1	0
chr7:54854750-54854806	ACACTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	0	4	0
chr2:6903549-6903605	ACGCTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	0	2	0
chr10:31567534-31567630	ACACTTTATTTTCATTAAGTTG	Intergenic	L1PA7	1	0	1	0
chr6:23431503-23431654	ACGCTTTATTTTCATTAAGTTG	Intergenic	L1PA8	1	0	1	0
chr12:88668820-88668884	ACACTTTATTTTCATTAAGTTG	Intergenic	L1PA7	6	1	0	0
chr10:27588726-27588885	CCACTTTATTTTCATTAAGTTG	Intergenic	L1PA7	3	2	0	0
chr12:10106161-101061700	ACACTTTATTTTCATTAAGTTG	Intron of ANO4 gene	L1PA7	1	7	6	0
chr1:183683917-183683997	ACACTTTATTTTCATTAAGTTG	Intron of RGL1 gene	L1PA7	2	1	1	0
chr2:58169734-58169833	ACACTTTATTTTCATTAAGTTG	Intron of FANCL gene	L1P2	2	50	1	0
chr2:224486485-224486529	ACGCTTTATTTTCATTAAGTTG	Intron of CUL3 gene	L1P3	2	125	0	0
chr3:169414717-169414803	ACACTTTATTTTCATTAAGTTG	Intron of MECOM gene	L1P3	2	50	0	0
chr17:35525040-35525120	TTGCTTTATTTTCATTAAGTTG	Intron of SLFN12L gene	L1PA7	3	9	0	0
chr14:54406591-54406666	ACACTTTATTTTCATTAAGTTG	Intron of CDKN3 gene	L1P2	1	9	0	0
chr3:66311241-66311321	ACACTTTATTTTCATTAAGTTG	Intron of SLC25A26 gene	L1PA8	2	4	0	0

chr1:113574938-113575026	ATGCTTTATTTTCATTAAGTTG	Intron of MAGI3 gene	L1PA8	2	2	0	0
chr20:13053749-13053819	ACGCTTTATTTTCATTAAGTTG	Intron of SPTLC3 gene	L1P3	0	1	1	0
chrX:44254483-44254552	ACGCTTTATTTTCATTAAGTTG	Intron of EFHC2 gene	L1PA7	1	0	1	0
chr2:159357384-159357546	ACGCTTTATTTTCATTAAGTTG	Intron of BAZ2B gene	L1PA7	0	0	5	2

	Hotspots in hESCs
	Hotspots in HT1080
	Hotspots in both HT1080 and hESCs

3.2. λ -Integrase mediated seamless vector transgenesis platform for biologics production

3.2.1. Strategy for biologics production platform in HEK293T

Besides utilizing λ -integrase mediated transgenesis tool for expression of a therapeutic protein, another application of this tool in the field of biologics production was also explored. The strategy to develop that λ -integrase mediated biologics production platform involved the following: (a) generation of a master cell line by random insertion of landing pads carrying λ -integrase target site, *attH4X* sequence (b) λ -integrase mediated targeting of the landing pad with a seamless vector carrying a promoterless reporter gene (c) primary screening of the master cell line based on the high expression of the reporter gene from the landing pad leading to preselection of the transcriptionally active sites for biologics production (d) integration of transgene encoding therapeutic protein at the landing pads for its production and testing.

For this purpose, a plasmid, *pPGKss-4X-attH4X-Hygro*, comprising of a promoter (PGK) followed by four copies of *attH4X* sequences along with a hygromycin resistant gene was designed. Four copies of *attH4X* sequences were used in the plasmid to increase the local concentration of *attH4X* recombination site at a specific genomic locus to enhance recombination efficiency by Int-C3. The plasmid, *pPGKss-4X-attH4X-Hygro*, was digested with restriction enzymes *ClaI* and *NotI*, to generate a linearized vector, termed as a landing pad, for random integration in HEK293T cells (figure 3.28 a). HEK293T cells were transfected with 300ng of the linear landing pad, followed by hygromycin selection for 1.5 weeks which resulted in 93 docking clones. The clones were grown on a 96-well plate and subsequently split into two plates. One plate was maintained as parental clones

and the other was used to co-transfect with a 3.8kb seamless target vector, *attL4X-Puro-EF-GFP* and Int C3 expression plasmid (figure 3.28 b). The supercoiled seamless vector *attL4X-Puro-EF-GFP* was generated from *pattP4X-Puro-EF-GFP-attH4X* through *in vitro* recombination. The idea is that the transfection of seamless target vector with Int-C3 in the clones harbouring landing pads will result in the integration of seamless vector at the *attH4X* sites of the landing pads. Due to this integration, the puromycin resistant gene will be expressed from *PGK* promoter of landing pads, thus, puromycin resistance is added to the cells, which still remain hygromycin-resistant (figure 3.28 a).

Figure 3.28. Strategy for biologics production platform in HEK293T cells. a) A schematic illustration showing random insertion of linear landing pad (*PGKss-4X-attH4X-Hygro*) in HEK293T genome. This was followed by the targeting of one of the *attH4X* sites in the landing pad with seamless target vector, *attL4X-Puro-EF-GFP* by Int-C3, leading to the integration of the Puro-EF-GFP cassette in between PGK and Hygromycin cassette. b) Experimental strategy demonstrating the landing pad insertion and λ -Int-mediated targeting of seamless vectors in the landing pads of HEK293T cells. This includes transfection of the linear landing pad in HEK293T cells followed by selection and isolation of hygromycin resistant clones. The clones were replicated into two plates and one of the plates was used for co-transfection of seamless target vectors and Int-expressing plasmid. This was followed by selection, expansion, and characterization of the targeted clones using PCR, sequencing and functional studies and the successfully tested clones were selected from parental clone plate for downstream applications.

3.2.2. Targeting of HEK docking clones with seamless vector *attL4X-Puro-EF-GFP*

The docking clones were co-transfected with 20ng of seamless target vector and 20ng of Int-C3 expression plasmid on a 96 well plate. After 2 weeks of puromycin selection, puromycin resistant and EGFP⁺ clones were obtained. Eighteen clones out of 93 clones were selected based on high expression of EGFP and expanded for PCR analysis. Genomic DNA from wild-type HEK 293T cells and parental HEK docking clones were used as negative controls for PCR analysis. At first, the PCR analysis was carried out by using primers pairs that bind to the landing pad for complete amplification of recombinant products at the landing pad (top panel of figure 3.29 a). However, this PCR analysis resulted in a product of size 4.2kb which was detected in only two clones, #3T and #17T (figure 3.29 a). The rest of the clones showed a PCR product of 400bp corresponding to amplification of untargeted landing pad (figure 3.29 a). Thus, second PCR analysis was carried out by using primers pairs that bind to landing pad and target vector for specific amplification of recombinant junctions (top panel of figure 3.29 b, c, d). These PCRs confirmed the recombinant left and right junctions generated by seamless target vector insertion at the landing pad for 9 clones (figure 3.29 b, c, d). No PCR products for junctions were detected for genomic DNA obtained from wildtype HEK 293T cells and parental HEK docking clones. Sequencing analysis of PCR products from left and right junctions for three clones (181T, #11T and #18T) confirmed that each clone had at least one seamless target vector integration in one of the *attH4X* sequences of the landing pads (table 14). It suggested that all four *attH4X* sequences of the landing pads could be accessible for the λ -Int mediated site-specific integration of the seamless vector.

(b)

PGK Fwd + GFP Rev 151

GFP Fwd 599 + Hygro rev

Figure 3.29. Screening for seamless vector integration at *attH4X* site of landing pads in HEK293T cells. PCR analysis using the genomic DNA from puromycin resistant clones obtained from targeting of seamless target vector *attL4X-Puro-EF-GFP* in HEK docking clones. Position of primers PGK fwd, Hygro rev, Puro rev 102, GFP fwd 599 and GFP rev 151 to identify recombinant junctions after successful integration at *attH4X* in landing pads are indicated. a) PCR screening with PGK fwd and Hygro rev (~4200bp) resulted in 2 positive clones. b) PCR-amplifications with PGK fwd and Puro rev 102 (~850bp) resulted in 9 positive clones. c) PCR with GFP fwd 599 and Hygro rev (~1000bp) detected for 5 clones. d) PCR screening with PGK fwd and GFP rev 151 (~2700bp) resulted in 9 clones. W, no DNA template control; WT, genomic DNA from HEK293T wild-type cells; UN, genomic DNA from hygromycin resistant parental HEK docking clones obtained through transfection of 300ng of *PGKss-4X-attH4X-Hygro*; 1-20, genomic DNA from puromycin resistant clones obtained through co-transfection of 20ng of *attL4X-Puro-EF-GFP* and 20ng of *pEFssInt-C3CNLS*; L, 1kb DNA ladder.

Table 14. Location of *attH4X* sites of the landing pads of targeted HEK docking clones identified using different primer pairs shown in figure 3.29 b, c, d.

Clones	Targeted <i>attH4X</i> site of the landing pad
#3T	2 nd
#11T	1 st
#18T	1 st

3.2.3. FACS analysis for EGFP expression of targeted HEK docking clones

EGFP expression of targeted HEK docking clones was analyzed by flow cytometry. The FACS histogram clearly distinguishes the EGFP expressing population from parental HEK docking clones. FACS histogram of the EGFP population for three clones #3T, #11T, and #18T showed 92.5%, 99% and 87% of EGFP⁺ cells respectively (figure 3.30). A single peak for targeted HEK clones in the FACS histogram suggested homogenous EGFP expression for all three clones.

Figure 3.30. EGFP expression of targeted HEK docking clones. Flow cytometric analysis of EGFP expression of targeted HEK docking clones #3T, #11T, and #18T. A single parameter histogram overlay plots representation of parental HEK docking clones onto the targeted HEK docking clones resulting from co-transfection of seamless target vector, *attL4X-Puro-EF-GFP* with Int-C3 expression plasmid; clone#3T (left panel), clone#11T (middle panel) and clone#18T (right panel).

3.2.4. Southern Blot analysis of parental and targeted HEK docking clones

To detect the copy number of landing pads integration in the HEK293T genome, the parental HEK docking clones (#3, #11 and #18) were analysed by Southern blotting. The genomic DNA from the clones were digested with *XbaI* and hybridized to probes complementary to *Hygro*. The Southern blot analysis of parental HEK docking clone #3, #11 and #18 revealed that the linear landing pad had

formed concatemers which might have been inserted at a single genomic locus for clone #3 and #18 and two genomic loci for clone #11 (figure 3.31 a). Since *XbaI* cleaves only once (downstream of *Hygro* gene) in the landing pad, the 2.3kb bands in the blot for all the clones represent unit size length of linear landing pads. Any other bands except 2.3kb represent the genomic location for the landing pad insertion. The *XbaI* site is also present in the HEK293T genome and thus, digestion of the genomic DNA from the clones with *XbaI* will result in linear DNA fragments of size >1.9kb that corresponds to the leftmost landing pad along with genomic DNA. This linear DNA upon hybridization with probes complementary to *Hygro* will show a band of size >1.9kb. For clone #3, there was only one band (>10kb) present in the southern blot except for 2.3kb band that suggested the landing pad had formed concatemers which might have been inserted at a single genomic locus. In case of clone #18, besides 2.3kb band, there was another band at ~4kb, suggesting the insertion of the landing pads at a single genomic locus. Similarly, for clone #11, the landing pad concatemers might be inserted in two genomic loci as suggested by two DNA bands (~4kb and ~2.1kb) in the southern blot.

Additionally, the targeted HEK docking clone #3T was analysed by Southern blot to detect the copy number of seamless target vector being integrated at landing pads. The genomic DNA from the clone was digested with *ScaI* and labelled using the probes complementary to *Hygro* and *Puro* gene. The southern blot analysis of the clone #3T using two different probes revealed that the seamless target vector was integrated as a single copy at *attH4X* site of the landing pad (left and middle panel of figure 3.31 b). To further detect the position of the landing pad that had been targeted with seamless vector, the genomic DNA from clone #3T was digested with *BsrGI* and labelled using the probes complementary to *EGFP*. The blot revealed that the seamless vector was targeted at the left-most landing pad of the concatemer in the parental clone #3 (right panel of figure 3.31 b).

Probe: Hygro gene

Probe: Hygro gene

Probe: Puro gene

Probe: EGFP gene

Figure 3.31. Southern blot analysis of parental and targeted HEK docking clones. a) A southern blot analysis for parental HEK docking clones showed the expected size bands (indicated by arrows), obtained using PCR-derived digoxigenin-labelled probe complementary to *Hygro*. The southern blot was carried out using genomic DNA purified from hygromycin resistant parental HEK docking clones digested with *XbaI*. b) Southern blot analysis for targeted HEK docking clone #3T showed the expected size bands (indicated by arrows) obtained using PCR-derived digoxigenin-labelled probe complementary to *Hygro* (left panel), *Puro* (middle panel) and *EGFP* (right panel) gene. The southern blot was carried out using with genomic DNA purified from targeted HEK docking clone #3T digested with *ScaI* and *BsrGI*. Lanes: landing pad, 10^8 , 10^7 copies of linearized landing pad, *PGKss-4X-attH4X-Hygro* loaded as positive control; pTV, 10^8 copies of *pattP4X-Puro-EF-GFP-attH4X* loaded as positive control, WT, genomic DNA from HEK293T wild-type cells; Doc #3, #11, #18, genomic DNA obtained from parental HEK docking clone #3, #11 and #18 (hygromycin resistant clones obtained from transfection of 300ng of *PGKss-4X-attH4X-Hygro* in HEK293T); #3T, genomic DNA from targeted HEK docking clone #3 (a puromycin resistant clone obtained from co-transfection of *attL4X-Puro-EF-GFP* and plasmid expressing Int-C3 in HEK docking clone #3).

3.2.5. Application of seamless vector transgenesis platform for factor VIII protein production

The biologics production platform in HEK293T was further utilized to integrate full-length factor 8 (FLF8) gene and express factor VIII protein, to test its utility in the biopharmaceuticals production sector. Factor VIII (FVIII) is a vital blood-clotting protein that is encoded by the *F8* gene in humans. Mutations in the *F8* gene are known to cause hemophilia A, a recessive X-linked coagulation disorder. The mutations lead to either low or abnormal production of factor VIII protein. This modified protein cannot perform efficiently in the blood clotting procedure leading to blood loss (304). The main treatment of hemophilia A is replacement therapy that involves the intravenous injection of factor VIII protein to replace the clotting factor that is missing or low in the bloodstream (305). Even though the development of recombinant factor VIII products was successful in the treatment of hemophilia A, unfortunately only one-third of the patients managed to avail recombinant factor VIII products due to extremely high product cost and inefficient commercial manufacture (306,307). Hence, there is a high demand for factor VIII protein production in the biopharmaceuticals sector as a cost-effective and efficient approach to treat hemophilia A patients.

Recombinant Factor VIII protein is mainly produced in large scale using several mammalian cell lines, such as Chinese Hamster Ovary cells (CHO), Baby Hamster Kidney (BHK), Mouse myeloma NS0 and SP2/0 derived cells and Human Embryonic Kidney cells (HEK293-T) (308). Several commercially available recombinant factor VIII products such as Recombinate®, Kogenate® FS, Helixate® FS, Advate® are approved by FDA for Hemophilia A treatment. These products are developed by genetically modifying CHO, BHK and HEK293 cell lines using tools such as lentiviral, adenoviral, transposon vectors, to randomly insert the full-length FLF8 gene for the recombinant protein production (308,309). One of the major bottlenecks for production of recombinant factor VIII protein is the large size of factor VIII cDNA (7kb) which complicates its delivery and insertion in the mammalian cells. The full-length F8 gene is one of the largest genes of size ~180kb that comprises of 26 exons and 25 introns. It is transcribed into 9kb of exon and 177kb of introns that encodes a polypeptide chain of 2351 amino acids (310,311).

It is due to its large size that only a few types of viral vectors are suitable for its packaging and delivery. Additionally, upon packing large sized transgene in viral particle, the production of an optimal number of effective viral particles for transduction is drastically reduced. Thus, viral vectors are inefficient for biologics production with large transgenes (312). Even though retroviral and adenoviral vectors are not constrained by the transgene size, it has been reported that the inserting >7kb transgene reduces the generation of effective viral particles and also the administration of recombinant proteins from viral particles elicits immune responses in the host cells (313,314). On the other hand, packaging of a transgene of size >9kb in adeno-associated viral (AAV) vectors had resulted in the generation of defective viral particles and thus encapsulating incomplete transgene sequence and drastically reducing the transduction efficiency (315,316). Even though lentiviral vector can insert transgene up to size 7kb, yet its application with respect to large transgene is limited due to reduction in the viral particle production with an increase in insert size which has been predominantly observed while packaging 4.5kb B domain deleted Factor VIII transgene in the virus (317-319). It is due to the large size of human full-length factor VIII transgene, most of the gene therapy studies are now focused in manipulating the transgene construct by introducing splice donors and acceptor sites and deleting the coding sequence that is unnecessary for protein expression and function (312). In this respect, the full-length factor VIII transgene of size 7kb was truncated by removing non-essential B-domain, thus reducing the size to 4.5kb (termed as BDD Factor VIII) (320). Non-viral vectors have also been explored for large transgene delivery as they provide a window to overcome immunotoxicity and transgene size limitation (312). An example of the most commonly used non-viral vectors is transposon elements. Studies had shown that the *Sleeping Beauty* transposon had been successfully used to insert BDD Factor VIII in the hemophilia A mouse model (321) and the piggyBac transposon could insert up to 9kb transgene with a moderate integration efficiency in mammalian cells (322). The piggyBac transposon was also applied to insert full-length factor VIII transgene in cell lines and mouse models to successfully rescue bleeding phenotype from hemophilia A mouse model (323). Although transposons can deliver large transgenes, their application is quite limited due to lower integration efficiency and the transgene size is constrained within 10kb (324). In this respect, λ -integrase can act as a powerful tool to site-specific insertion of the large transgene for recombinant factor VIII production. It is well known that during the lysogenic life cycle of phage λ , wild-type λ -integrase can integrate its complete genome of size 48kb site-specifically at the bacterial genome with high efficiency (164). Hence, as a proof of concept, the next focus of the study was to utilize the λ -integrase mediated transgenesis platform for insertion of a large transgene, such as full-length F8 gene, at the landing pads of the master cell line to obtain stable expression of factor VIII protein.

3.2.6. Strategy for FVIII protein production platform in HEK293T

To develop therapeutic FVIII protein production platform, an FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*, of 14.2kb in size was generated that composed a promoterless puromycin resistant cassette along with full-length F8 (FLF8) cDNA and a fluorescence marker ZsGreen under EF promoter (figure 3.32). Parental HEK docking clone #3 was co-transfected with two different amounts (500ng and 1µg) of FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* and expression plasmids for either Int-C3 or inactive Int (1µg). After 2 weeks of Puromycin selection, bulk cultures were expanded for each amount of transfected plasmid for extraction of genomic DNA to carry out PCR/sequence analysis.

Figure 3.32. Strategy for therapeutic protein FVIII production platform in HEK293T. A schematic illustration showing random insertion of linear landing pad (*PGKss-4X-attH4X-Hygro*) in HEK293T genome. This was followed by the targeting of one of the *attH4X* sites in the landing pad with FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* by Int-C3, leading to the integration of the Puro-EF-FLF8-IRES-ZsGreen cassette in between PGK and Hygromycin cassette.

3.2.7. PCR Screening for HEK docking bulk cultures targeted with *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*

PCR analysis for the junctions generated by plasmid integration at the landing pad was performed using primer pairs in landing pads and the plasmid vector as shown in figure 3.33 a. As studied in section 3.1.9 that transfection of substrate plasmid might lead to insertion of the seamless vector (generated by *ex vivo* intramolecular recombination) at the target site. Thus, PCR analysis was carried out for the unique right junction generated by seamless vector integration at the landing pad using primer pairs as shown in figure 3.33 b. However, PCR products were only detected for junctions generated by plasmid integration at the landing pad with genomic DNA obtained from the targeted bulk culture of 1µg of FLF8 vector with Int-C3 (figure 3.33 c, d, e). This suggested that the 14.2kb

plasmid, not the seamless vector, was integrated at the *attH4X* site of the landing pad by Int-C3. The sequence analysis of PCR products from left and right junctions confirmed targeting of FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*, at first *attH4X* site of the landing pads of HEK docking clone #3. No desired PCR product was obtained with genomic DNA obtained from the targeted bulk culture of 500ng of FLF8 vector with Int-C3.

Figure 3.33. PCR analysis for *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* integration at *attH4X* site of landing pads in HEK293T cells. (a) A schematic representation of insertion of FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* by Int-C3 at the landing pad. Position of primers PGK fwd, Hygro rev, Puro rev 103, F8 rev and PR 21 to identify recombinant junctions after successful integration at *attH4X* in landing pads are indicated. (b) A schematic representation of insertion of FLF8 seamless vector, *pattL4X-Puro-EF-FLF8-IRES-ZsGreen* by Int-C3 at the landing pad. Position of primers PGK fwd, Hygro rev, Puro rev 103, and ZsGreen Fwd to identify recombinant junctions after successful integration at *attH4X* in landing pads are indicated. PCR analysis revealing the integration of *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* at *attH4X* site in landing pads of parental HEK docking clone #3. PCR was performed for left junction with primers PGK fwd and Puro rev 303/F8 rev (c,d) and for right junction with primers PR 21 fwd and Hygro rev (e) using the genomic DNA from two independent transfections with 500ng and 1000ng of *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* with 1000ng of Int-C3/Inactive Int in HEK docking clone #3. PCR amplified products of the expected size for the left junction (~650bp and ~3000bp) and the right junction (~500bp) were detected in bulk C3-2. L, 1kb DNA ladder W, no DNA template control; Doc 3, genomic DNA parental HEK docking clone #3. IN1/IN2: genomic DNA from puromycin resistant bulk culture obtained through co-transfection of 500ng and 1000ng of *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* and 1 μ g of *pEFssInt-Ina* in parental HEK docking clone #3 respectively. C3-1, C3-2: genomic DNA from puromycin resistant bulk culture obtained through co-transfection of 500ng and 1000ng *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* and 1 μ g of *pEFssInt-C3CNLS* respectively.

3.2.8. FACS sorting and FVIII protein activity test for HEK docking bulk cultures targeted with *pattP4X-Puro-EF-F8-IRES-ZsGreen-attH4X*

After confirming the integration of the FVIII expressing plasmid, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*, at the landing pads of HEK docking clone #3, the next focus of the study was to

analyse the transgene expression in terms of ZsGreen⁺ cells and FLF8 protein activity. The FACS analysis of the bulk culture revealed that only 5.6% of the bulk culture was having ZsGreen expression (figure 3.34 a, left panel). Hence, the 5.6% of ZsGreen⁺ cells were sorted by FACS and the percentage of the ZsGreen⁺ cells was enriched to 85% as shown in figure 3.34 a (right panel). The PCR analysis on the sorted cells further confirmed that the sorted ZsGreen⁺ cells harboured FLF8 vector at the landing pads of HEK docking clone #3 (data not shown).

In addition, the sorted bulk culture was also tested for FVIII protein activity by a chromogenic assay. This assay kit is based on the ability of activated factor VIII (FVIIIa) to generate activated factor X (FXa). The secreted factor VIII (FVIII) protein from HEK293 cells is activated in the presence of calcium ions, thrombin and negatively charged membrane phospholipids (provided in the kit). The activated factor VIII (FVIIIa) protein then binds to the activated Factor IX (FIXa). This complex then proteolytically activates factor X (FX) to factor Xa (FXa). The generated FXa proteolytically cleaves a specific Factor Xa chromogenic substrate (SXA-11, provided in the kit) and releases para-nitro aniline (*pNA*) chromophore, which is quantified from colour development at 405nm using a microplate reader.

The assay data were normalized to cell viability (obtained from MTT assay) to calculate the fold change of protein activity over parental HEK docking clone. The chromogenic assay revealed that the fold change in FVIII protein activity from the bulk culture was only ~3.5 times higher than the parental HEK docking clone #3 whereas the sorted bulk culture showed a fold change of ~20 times for factor VIII protein activity over parental clone (figure 3.34 b). Based on the standard curve of FVIIIa Enzyme as shown in figure 3.34 c, the total yield of FVIII protein from the sorted bulk culture was estimated to be ~2.27µg/ml (table 15) as compared to the unsorted bulk culture that showed a yield of only 0.03µg/ml of factor VIII protein. This further suggests high FVIII protein expression from integrated FLF8 vector at the landing pads.

Figure 3.34. ZsGreen and factor VIII expression from HEK bulk culture targeted with FLF8 vector. a) ZsGreen expression of HEK bulk culture targeted with FLF8 vector. Left panel: a single parameter histogram overlay plots representation of parental HEK docking clone #3 onto the targeted HEK bulk culture obtained from co-transfection of FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen attH4X*, with Int-C3 expression plasmid. Right panel: an overlay obtained with targeted HEK bulk culture and its sorted ZsGreen⁺ cells showed enrichment of ZsGreen⁺ cells after FACS sorting. (b) Chromogenic assay to test the activity of secreted FVIII protein from the HEK targeted sorted culture over HEK Doc #3. The bar graph represents fold change in the activity of secreted FVIII protein after normalizing to cell viability through MTT assay. HEK Doc 3: supernatant from parental HEK docking clone #3; HEK targeted bulk culture: supernatant from the bulk culture obtained from co-transfection of 1000ng of *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* and 1 μ g of *pEFssInt-C3CNLS*; HEK targeted sorted culture: supernatant from the FACS sorted ZsGreen⁺ cells obtained from HEK targeted bulk culture. (c) Standard curve of FVIIIa Enzyme obtained by factor VIII activity assay kit.

Table 15. Estimated yield of factor VIII protein from HEK targeted bulk culture and HEK targeted sorted culture based on the factor VIII protein standard curve in figure 3.34 c

Samples	Absorbance at 405nm	Final absorbance (A405- control)	Amount of factor VIII protein (ng)/5µl	Amount of factor VIII protein (µg/ml)
DMEM media	0.14	0.00	0	0
HEK Doc #3	0.14	0.00	0	0
HEK targeted bulk culture	0.16	0.02	0.11	0.03
HEK targeted sorted culture	1.74	1.60	11.34	2.27

3.2.9. Attempts to obtain clones for HEK docking clones targeted with *attL4X-Puro-EF-F8-IRES-ZsGreen*

In order to obtain the clonal population that harbours FLF8 vector at the landing pads, parental HEK docking clone #3 was first co-transfected with two different amounts (50ng and 1µg) of 11kb seamless vector, *attL4X-Puro-EF-F8-IRES-ZsGreen*, and expression plasmid Int-C3 (1µg) or Inactive-Int (1µg). The seamless vector, *attL4X-Puro-EF-F8-IRES-ZsGreen*, was generated from *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* through *in vitro* intramolecular recombination. The number of clones obtained for each amount of transfected seamless vector after puromycin selection is listed in table 16. The clones were first screened for secreted FVIII protein activity and followed by PCR analysis. The two clones obtained from co-transfection of 50ng of seamless vector and Int-C3 did not show any significant levels of FVIII protein activity over parental clone in the chromogenic assay (table 17). However, all the 18 clones obtained from co-transfection of 1µg of seamless vector and Int-C3 showed 2-3 times higher factor VIII activity over the parental clone (table 17). These 18 clones were further analysed with PCR using primers as shown in figure 3.33 b to amplify recombinant junction generated by the seamless vector integration at landing pads. Surprisingly, none of the clones showed desired PCR amplification for recombinant junctions. Southern blot analysis on these 18 clones with a probe complementary to *Hygro* further confirmed non-integration of the seamless vector at the landing pad (data not shown).

Another round of targeting was performed by co-transfecting 1µg of FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X*, with expression plasmid Int-C3 (1µg) or Inactive-Int (1µg) (table 16). After 2 weeks of puromycin selection, 22 clones were obtained from co-transfection of 1µg of FLF8 vector and Int-C3. These clones were screened for FVIII protein activity; out of which 18 clones showed 3-4 times higher activity over the parental clone (table 17). However, PCR analysis using primers as shown in figure 3.33 a for these 13 clones did not show any desired products for the left and right junction of FLF8 vector integration at landing pads. Southern blot analysis on these 18 clones with a probe complementary to *Hygro* confirmed non-integration of FLF8 vector at the landing pad (data not shown).

Table 16. Summary of targeting events for FLF8 vector transgenesis in HEK docking clone #3.

FLF8 vector	Int expression plasmid	No. of clones obtained	No. of clones having FVIII activity
<i>attL4X-Puro-EF-F8-IRES-ZsGreen</i> (50ng)	Int-C3	2	0
<i>attL4X-Puro-EF-F8-IRES-ZsGreen</i> (50ng)	Inactive-Int	1	0
<i>attL4X-Puro-EF-F8-IRES-ZsGreen</i> (1µg)	Int-C3	18	18
<i>attL4X-Puro-EF-F8-IRES-ZsGreen</i> (1µg)	Inactive-Int	0	0
<i>pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X</i> (1µg)	Int-C3	22	13
<i>pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X</i> (1µg)	Inactive-Int	Bulk culture	0

Table 17. Summary of fold change for secreted FVIII protein activity from clones obtained from FLF8 vector transgenesis in HEK docking clone #3.

<i>attL4X-Puro-EF-F8-IRES-ZsGreen</i> (50ng)	
Clones	Fold change for FVIII protein activity
HEK Doc 3	1
Inactive clone	0.86
A1	0.79
A2	0.88
<i>attL4X-Puro-EF-F8-IRES-ZsGreen</i> (1µg)	
Clones	Fold change for FVIII protein activity
HEK Doc 3	1.00
#1	2.13
#2	3.05
#3	3.22
#4	2.66
#5	2.82
#6	1.84
#7	2.80
#8	1.83
#9	3.11
#10	3.05
#11	2.20
#12	2.98
#13	2.55
#14	2.16
#15	3.28
#16	3.28
#17	2.93
#18	3.32
<i>pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X</i> (1µg)	
Clones	Fold change for FVIII protein activity

HEK doc 3	1
Inactive bulk	0.63
#1	0.06
#2	0.09
#3	0.03
#4	3.28
#5	3.74
#6	3.77
#7	3.33
#8	3.48
#9	3.76
#10	0.11
#11	0.06
#12	4.15
#13	4.18
#14	4.32
#15	4.06
#16	4.16
#17	0.07
#18	0.08
#19	3.41
#20	0.07
#21	3.79
#22	0.06

3.3. Enhancement of transgene expression from site-specifically targeted seamless vector carrying CTCF binding motifs

3.3.1. CCCTC-binding factor (CTCF)

After exploring two different applications of λ -integrase transgenesis platform, the next aim of the study was to improve transgene expression by using a DNA element known as the CCCTC-binding factor (CTCF). CTCF is a zinc-finger protein composed of an N-terminal domain, a central zinc-finger domain with 11 zinc fingers (ZF) and a C-terminal domain as shown in figure 3.35 (325,326). CTCF binds to 15bp core motif of DNA using ZFs 3-7 and remaining ZFs domains modulates the stability of the complex by interacting with adjacent DNA molecules. All the domains of CTCF also interact with different proteins, RNA and can also undergo post-translational modifications (325). CTCF widely binds to 40,000-80,000 sites in the mammalian genome which overlaps with introns, intergenic and regulatory sites like enhancers and promoters (327). Studies have shown that CTCF is associated with the formation of chromatin loops that are considered as topological structures inside which transcription takes place (261,328).

Figure 3.35. Domain structure of CTCF: the three major domains: N-terminal domain, central zinc-finger domain (containing Zn-fingers 1-11) and C-terminal domain [reprinted with permission (327)].

The eukaryotic genome is organized in multiple scales where each scale interplays between genome structure and its functionality as shown in figure 3.36. The smallest scale represents nucleosome wrapping 147bp DNA that functions to compact DNA structurally into chromatin and regulates gene expression by epigenetic modifications. The intermediate scale ranges from kilobases to megabases where the chromatin is further organized into three-dimensional domains. The first type of domain is known as Topological Associating Domains (TADs) that regulate promoter-enhancer activity by using architectural proteins such as cohesin and CTCF for the formation of chromatin loops. The second type of domain is called A and B compartments that represent “active” and “inactive” state of chromatin (figure 3.36). Both these domains are formed by a distinct mechanism where TADs represent the epigenomic compartments by establishing intra and inter-chromosomal contacts between other domains. Lastly, at the largest scale, chromosomes are tightly packed into chromosomal territories positioned into specific locations inside the nucleus (329-331).

Figure 3.36. Hierarchy of chromosome structures: nucleosomes, TADs, epigenomic compartments and chromosome territories [reprinted with permission (331)].

Several Hi-C studies had noticed that the TADs boundaries are universally enriched with cohesin and CTCF which suggests that both of them plays an important role in TADs formation and maintenance (261,328). Cohesin composed of Smc1, Smc3, SA1/2 and Rad21 that forms a ring-shaped multiplex structure which can topologically entrap DNA (332) (figure 3.37 a). It can also move on both naked DNA and chromatin (333) which leads to the formation of TADs in the presence of CTCF site with a tracking mechanism called the “loop extrusion model”. In this model, cohesin molecule spontaneously slides on the chromatin and extrudes a DNA loop. The DNA loop extrusion continues until cohesin encounters CTCF occupied convergently oriented sites on each arm of the loop. The loop extrusion process then stops and CTCF comigrates with cohesion (334,335) (figure 3.37 b). Most of the CTCF binding sites in the genome are found in convergent orientation (325,336,337). Several studies had shown the CTCF-cohesin complex is dynamic in nature where the DNA loops continuously break and reform throughout the cell cycle process (338).

Figure 3.37. Cohesin structure and DNA loop extrusion model. a) Structure of cohesion comprised of Smc1, Smc3, Rad21 and SA1/2 that entraps the DNA strand (338). b) DNA loop extrusion model by CTCF and cohesion [reprinted with permission (338)].

CTCF binding sites consist of three motifs as shown in figure 3.38 a. A set of nonpalindromic CTCF-binding sites with a sequence consensus is considered as the DNA binding domain, termed as M1. It is a 20-base-pair (bp) core motif that is known to be identified by most of the ChIP seq experiments (figure 3.38 b). This site engages zinc fingers 4–7 for stable DNA binding. Another 10bp motif, termed M2, is found upstream of M1 in 15%– 25% of the CTCF sites and it helps to stabilize CTCF overall binding by engaging zinc fingers 9–11. A 10bp motif is found downstream to M1, which is known to destabilize CTCF binding by the disengagement of zinc fingers 1–2 (325). CTCF is thought to affect local genomic methylation by excluding DNA methyltransferases from the region, thus regulating gene expression. CTCF binding itself appears to be sensitive to methylation, so a lack of CTCF binding can lead to default methylation (339). Moreover, CTCF can influence DNA methylation by forming a complex with two enzymes related to DNA methylation: poly(ADP-ribose) polymerase 1 (PARP1) and the ubiquitously expressed DNA (cytosine-5)-methyltransferase 1

(DNMT1) (340). This would lead to inhibiting transgene silencing by CTCF mediated exclusion of DNA methyltransferase at its binding sites (212).

Figure 3.38. CTCF binding sites. a) CTCF-binding motifs showing the M1/core and the M2/upstream and downstream sequences (325) b) DNA logo of M1 for CTCF-binding site obtained from several ChIP-on-chip experiments [reprinted with permission (336)].

3.3.2. The hypothesis of inclusion of the CTCF sites in the seamless vector for enhanced transgene expression

Upon considering all these properties of CTCF sites, the hypothesis was that the inclusion of CTCF binding sites in the seamless vector might result in enhanced transgene expression from the human genome by the formation of DNA loops and thus increasing the enhancer-promoter activity. For this study, the *UTF1* reporter construct, UE, (341) that comprises of the *UTF1* promoter, *EGFP* cassette, polyA and enhancer sequences, was exploited (figure 3.39). The aim was to explore whether the inclusion of CTCF binding site flanking the *UTF1* reporter construct increases communication between *UTF1* promoter and enhancer and thus resulting in enhanced EGFP expression from *UTF1* promoter as compared to the negative control vector without CTCF binding sites.

Figure 3.39. UTF1 reporter construct (UE) comprising of the *UTF1* promoter, *EGFP* cassette, polyA and enhancer sequences [adapted from (341)].

3.3.3. Strategy to include CTCF site for enhanced transgene expression

To test this hypothesis, three different multi-cassette substrate plasmids were generated harbouring two copies of 14bp DNA binding motifs of CTCF (as per figure 3.38 b) in convergent orientation at two different positions in the substrate plasmid. The first plasmid, *pattP4X-scCTCF-UTF1-GFP-*

scCTCF-PGKssPuro-attH4X, is a negative control vector that contained 14bp scrambled CTCF sequence flanking the *UTF1-GFP* cassette. The second plasmid, *pattP4X-CTCF-UTF1-GFP-CTCF-PGKssPuro-attH4X*, contained 14bp CTCF sequence flanking the *UTF1-GFP* cassette while the third plasmid, *pattP4X-CTCF-UTF1-GFP-PGKssPuro-CTCF-attH4X*, contained 14bp CTCF sequence flanking both cassettes. Supercoiled seamless vectors (4.5kb) were generated by *in vitro* intramolecular recombination from respective plasmids that were used for transfection in human embryonic stem cells (hESCs) (figure 3.40).

Figure 3.40. Schematic representation of seamless vectors containing CTCF sequences. a) *attL4X-scCTCF-UTF1-GFP-scCTCF-PGKssPuro* contains scrambled CTCF sequence flanking the *UTF1-EGFP* cassette b) *attL4X-CTCF-UTF1-GFP-CTCF-PGKssPuro* contains CTCF sequence flanking the *UTF1-EGFP* cassette c) *attL4X-CTCF-UTF1-GFP-PGKssPuro-CTCF* contains CTCF sequence flanking the *PGK-Puro-UTF1-EGFP* cassette

hESCs were co-transfected with 100ng of each of the three seamless vectors with Int-C3 expression plasmid (1 μ g) and puromycin selection was carried out for 2 weeks. The selection resulted in low *EGFP* expressing clones. The low expression of EGFP in these clones is most likely due to the fact that the *UTF1* promoter is inherently a weak promoter (342). Thus, to enhance the fluorescence signal, the *EGFP* cassette of all three substrate plasmids were replaced with *mNeonGreen* cassette which is considered as a brighter alternative to *EGFP*. The newly generated substrate plasmids, *pattP4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro-attH4X*, *pattP4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro-attH4X* and *pattP4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF-attH4X*, were used to produce the corresponding supercoiled seamless vector by *in vitro* intramolecular recombination from respective plasmids (figure 3.41).

Figure 3.41. Schematic representation of seamless CTCF vectors having *EGFP* cassette replaced with *mNeon*. a) *attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro* contains scrambled CTCF sequence flanking the *UTF1-mNeon* cassette b) *attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro* contains CTCF sequence flanking the *UTF1-mNeon* cassette c) *attL4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF* contains CTCF sequence flanking the *PGK-Puro-UTF1-mNeon* cassette

3.3.4. FACS analysis for *mNeon* expression in targeted hES clones

hESCs were transfected with 100ng of each of the three seamless vectors with expression vector Int-C3 (1 μ g) and puromycin selection was carried out for 2 weeks. The clones obtained were

first screened for mNeon expression followed by PCR analysis for seamless vector insertion at the human *LINE-1 element*. The number of mNeon⁺ clones obtained from each seamless vector and their respective percentage of mNeon⁺ cells are summarized in table 18. There was a 2-fold difference in the number of mNeon⁺ clones obtained from the negative control vector and the vectors carrying CTCF sequences. The mNeon expression of each clone was further analysed by flow cytometry as shown in figure 3.42. FACS analysis revealed that the percentage of mNeon⁺ cells is higher in clones obtained from vectors carrying CTCF sites as compared to the negative control vector (table 18, figure 3.42).

Table 18. Summary of the number of mNeon⁺ clones and their corresponding mNeon⁺ cells (%) from hESCs targeted with seamless CTCF vectors and negative control vector.

No. of mNeon ⁺ clones / Total no. of clones	Clones#	mNeon ⁺ cells (%)
<i>attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro</i>		
3/40	D1	99.6
	B2	77.1
	A3	99.6
<i>attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro</i>		
6/39	D7	93.3
	C9	95.3
	G9	61.8
	E9	20
	A12	14.4
	G12	98.7
<i>attL4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF</i>		
13/40	B1	98.4
	C1	28.4
	E1	44
	E2	85.7
	F2	63.1
	A3	39.7
	H1	99.3
	D3	99.6
	A4	67
	C4	97
	E4	59
	G7	98
	H7	95

(a) attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro

(b) attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro

Figure 3.42. mNeon expression of hES clones targeted with seamless CTCF vectors and negative control vector. A single parameter histogram representing of mNeon expression of (a) clones #D1, #B2 and #A3 obtained from co-transfection of *attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro* with Int-C3 (b) clones #D7, #C9, #G9, #E9, #A12 and #G12 obtained from co-transfection of *attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro* with Int-C3 (c) clones #B1, #C1, #E1, #E2, #F2, #A3, #H1, #D3, #A4, #C4, #E4, #G7, #E7 obtained from co-transfection of *attL4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF* with Int-C3. WT hESCs: non-transfected control. The percentage of mNeon⁺ cells mentioned in the right corner of each histogram.

3.3.5. PCR analysis for seamless vector insertion at *LINE-1* elements

The mNeon⁺ clones obtained from all three seamless vectors were expanded and screened for insertion into *LINE-1* elements by using genomic PCR with four established diagnostic primer pairs that were designed previously to bind the consensus ORF1 sequence of *LINE-1* elements (figure 3.43). However, PCR analysis for all the mNeon⁺ clones using consensus primers did not yield specific products for any of the junctions (data not shown).

Figure 3.43. Strategy for seamless vector transgenesis in human *LINE-1* elements. A schematic illustration showing the targeting of genomic *attH4X* site with seamless vector *attL4X-UTF1-mNeon-PGKssPuro* by Int-C3, leading to the integration of the *UTF1-mNeon-PGKssPuro* cassette. Positions of primers used to identify successful integration at *attH4X* sites in *LINE-1* elements are indicated.

To determine the genomic location of seamless vector transgenesis in the hESCs, genomic DNA from the mNeon⁺ clones was digested with enzyme *BsrGI*, self-circularized (to form a circular DNA) and then subjected to inverse PCR using oppositely oriented primers in the seamless vector as shown figure 3.44 a, b top panel. Importantly, only the clones having 90-100% mNeon expression were analysed. Inverse PCR resulted in the amplification of unique bands for left junction as shown in figure 3.44 a and right junction as shown in figure 3.44 b for each clone. These unique bands were absent from template obtained from wildtype hESCs. The sequencing analysis of the inverse PCR products revealed the genomic loci of the seamless vector integration for each clone as shown in table 19. Surprisingly, none of the genomic loci obtained through inverse PCR belongs to *LINE-1 elements*. Genomic location of three clones, D1, D3, C4 could not be determined as the seamless vector was inserted as concatemers in the genome.

Figure 3.44. Inverse PCR analysis for identification of the genomic location of seamless vector transgenesis. Schematic diagram showing *BsrGI* digestion of recombinant products of seamless vector transgenesis at *attH4X* site of hESCs. Positions of relevant primers used for inverse PCR to determine left and right junctions are indicated. Inverse PCR screening using the genomic DNA from puromycin resistant clones obtained from targeting of seamless CTCF vector and negative control vectors in hESCs. a) Recombinant left junction was identified involving semi-nested inverse PCR with primers 024_fwd and attP rev using genomic DNA template from primary PCR (with primers 029_fwd and attP rev). Inverse PCR amplification with 024_fwd and attP rev (>560bp) resulted in unique bands for 5 clones. b) The recombinant right junction was identified involving nested inverse PCR with primers fwd 509 and fwd 578 using genomic DNA template from primary PCR (with primers fwd128 and fwd 330). Inverse PCR amplification with fwd 509 and fwd 578 (>570bp) resulted in unique bands for 11 clones. W, no DNA template control; ES, genomic DNA from parental hESCs; #D1, #A3, puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro* and 1μg of *pEFssInt-C3CNLS*; #D7, #C9, #G12, puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro* and 1μg of *pEFssInt-C3CNLS*; #B1, #H1, #D3, #C4, #G7, #H7, puromycin resistant clones obtained through co-transfection of 100ng of *attL4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF* and 1μg of *pEFssInt-C3CNLS*, L, 1kb DNA ladder.

Table 19. List of genomic coordinates of hES clones targeted with seamless CTCF vector identified using inverse PCR.

Clones#	Genomic coordinates	Genomic Locations
<i>attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro</i>		
D1	Concatemer	
A3	chr10:29876816-29877781	Intergenic
<i>attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro</i>		
D7	chr18:48948983-48949806	Intron of SMAD7 gene
C9	chr10:91795570-91796081	Intron of TNKS2-AS1 gene
G12	chr2:203013416-203013867	Intergenic
<i>attL4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF</i>		
B1	chr12:121005742-121006172	Intron of C12orf43 gene
H1	chr22:42502467-42502616	Intron of SERHL gene
D3	Concatemer	
C4	Concatemer	
G7	chr12:101743025-101743855	Intergenic
H7	chr12:101743998-101744860	Intergenic

To further confirm the findings, the transfections of hESCs were repeated with each of the three seamless vectors and Int-C3 expression plasmid. Puromycin selection was carried out for 2 weeks and the obtained clones were first screened for mNeon expression followed by PCR analysis. The summary of the number of clones obtained from each seamless vector and their respective percentage of mNeon⁺ cells is shown in table 20. Unlike the last round of transfection, there was not much difference in the number of mNeon expressing clones between negative control and vectors carrying CTCF sequences. It might be due to the fact that the CTCF binding on DNA is sensitive to DNA methylation, thus, the epigenetic status of the genomic loci of seamless vector integration might affect CTCF binding to the seamless vector. Additionally, genomic PCR analysis using four established diagnostic primer pairs revealed that none of the clones harboured seamless vector at *attH4X* sequence of the human *LINE-1 elements* suggesting that the inclusion of CTCF sites in the seamless vector had somehow affected the λ -mediated seamless vector transgenesis into *LINE-1 elements*.

Table 20. Summary of targeting events for CTCF vector and negative control vector transgenesis in hESCs and their corresponding cell percentage for mNeon expression.

CLONES	Cells expressing mNeon (%)
<i>attL4X-scCTCF-UTF1-mNeon-scCTCF-PGKssPuro</i>	
1A	48
3C	82
3F	82
3G	100
4C	90
4H	17
5B	100
6D	45
7B	74
8D	100
8C	41
9G	100
9B	98
<i>attL4X-CTCF-UTF1-mNeon-CTCF-PGKssPuro</i>	
3G	75
3H	67
<i>attL4X-CTCF-UTF1-mNeon-PGKssPuro-CTCF</i>	
1H	100
2A	64
3E	21
4C	100
4E	53
5C	72
7B	74
8E	35
5D	40
6A	86
7F	50
9G	95

3.4. λ -Integrase mediated seamless vector transgenesis in actively transcribed *LINE-1* elements

3.4.1. Actively transcribed human *LINE-1* elements

To further increase the specificity of the λ -Integrase based transgenesis platform, the next aim was to identify any of the 933 *LINE-1* elements harboring *attH4X* sequences (1) are actively transcribed from an endogenous promoter in hESCs. *LINE-1* elements are a major group of retrotransposons that comprise ~21% of the human genome (343,344). About 500,000 *LINE-1* elements sculpted the human genome during human evolution. Full-length *LINE-1* element is about 6-7kb long that consists

of an untranslated 5'UTR followed by two open reading frames, ORF1 and ORF2, and 3'UTR region and flanked by ~9bp of target site duplication sites (343) as shown in figure 3.45.

Figure 3.45. Structure of *LINE-1* element having 5'UTR, ORF1 and ORF2, and 3'UTR region. TSD: Target site duplication [adapted with permission (345)].

The 5'UTR harbors an internal RNA Polymerase II transcription promoter in the sense orientation (343). The promoter activity is demonstrated by the first 100bp of the 5'UTR and this region completely lacks TATA-box for transcription initiation (346,347). The 5'UTR also consists of an antisense promoter which is recently identified as ORF0 of a length of 72 amino acids (348,349). The 5'UTR is subjected to CpG methylation which is believed to downregulate *LINE-1* transcription (343). ORF1 encodes for an RNA-binding protein required for ribonucleoprotein particle assembly and chaperone activity during retrotransposition. Most of the ORF1 sequences are non-functional due to the accumulation of random mutations and truncations acquired throughout human evolution (350,351). ORF2 encodes endonuclease and reverse-transcriptase enzyme that converts RNA to DNA for retrotransposition (343,344,352). 3'UTR region contains an internal weak polyA signal (AATAAA) that terminates the *LINE-1 elements* (353).

The process of retrotransposition starts by transcription of a polyadenylated mRNA from 5'UTR by RNA Polymerase II, followed by transport the mRNA to the cytoplasm and translation, and producing ORF1 (L1ORF1p) and ORF2 proteins (L1ORF2p) (figure 3.46). The mRNA gets bound by ORF1 and ORF2 proteins forming a ribonucleoprotein particle (L1-RNP) assembly which is transported to the nucleus where the retrotransposition occurs through target-site primed reverse transcription (TPRT) (354,355) (figure 3.46). Several transcriptional factors were studied to be associated with *LINE-1 element* transcription which includes T-cell factor/lymphoid enhancer factor (TCF/LEF), Yin Yang 1 (YY1), p53 and runt related transcription factor 3 (RUNX3) (343).

Figure 3.46. Retrotransposition cycle of *LINE-1* element. The steps include transcription, mRNA export, translation, ribonucleoprotein formation, import to the nucleus and target primed reverse transcription (reprinted from <https://www.pei.de/EN/research/groups/medical-biotechnology/medical-biotechnology-research-node.html>).

In human genome majority of *LINE-1* elements are transcriptionally inactive due to mutations, rearrangements or truncations at the 5' end and thus are incapable of retrotransposition. However, it was well studied that small subset of (80-100) of full-length *LINE-1* elements are functionally active and hence are capable of retrotransposition (356-358). *LINE-1* retrotransposition requires transcription of ORF1 and ORF2 of the *LINE-1* elements. There are few cell lines, such as HeLa, NTera2D1, HL60, hESCs and HEK293 where *LINE-1* RNA had been detected suggesting that some of the *LINE-1* elements might be actively transcribed (343,359). Two separate studies have also shown that there are about 147 intact full-length human *LINE-1* elements present in the human genome and only half of these elements were expected found to be retrotransposition competent. Most of these actively mobile *LINE-1* elements belongs to Ta-0, Ta-1d, Ta-1nd and pre-Ta subfamilies (356,360). Mills *et al.* also identified 80 *LINE-1* elements that have intact ORF2 sequence but lacks intact ORF1 (360). Chandra *et al.* recently showed the attH4X sequence (target site of λ -Integrase) is found in ORF1 of *LINE-1* elements at 933 different loci (1). Hence, the aim was to detect if any of the actively transcribed *LINE-1* elements harboring attH4X sequences can be targeted with λ -Integrase mediated seamless vector transgenesis tool.

3.4.2. Strategy for seamless vector transgenesis in actively transcribed *LINE-1*

To explore the possibility of targeting attH4X site in an actively transcribed *LINE-1* element in human embryonic stem cells (hESCs), a promoter-less plasmid (*patP4X-GFP-EF-Puro-IRES-attH4X*) was constructed that composed of IRES sequence followed by a Puromycin resistant gene and EGFP cassette expressed under EF1 α -promoter as shown in figure 3.47. It was used to produce the 4.2kb supercoiled seamless vector, *attL4X-GFP-EF-Puro-IRES*, through *in vitro* intramolecular

recombination. The seamless vector was designed in such a way that upon λ -Integrase mediated integration at *attH4X* sequence of actively transcribed *LINE-1* element, the IRES-Puromycin cassette will be in the same orientation of the 5'UTR promoter of *LINE-1* element. This will lead to expression of Puromycin resistant gene from the endogenous promoter of the *LINE-1* element (Figure 3.47).

Figure 3.47. Strategy for seamless vector transgenesis in actively transcribed human *LINE-1* elements. A schematic illustration showing the targeting of genomic *attH4X* site with seamless vector *attL4X-GFP-EF-Puro-IRES* by Int-C3, leading to the integration of the *IRES-Puro-EF-GFP* cassette under 5'UTR promoter of *LINE-1* element. Positions of primers used to identify successful integration at *attH4X* sites in *LINE-1* elements are indicated.

3.4.3. PCR screening for hES clones targeted with *attL4X-GFP-EF-Puro-IRES*

hESCs was co-transfected with 100ng of *attL4X-GFP-EF-Puro-IRES* and 1 μ g of Int-C3/Inactive-Int. Puromycin selection for 2 weeks yielded 4 clones with the seamless vector and Int-C3 and only 1 clone with the seamless vector and Inactive-Int. The clones were expanded to screen for insertion of the seamless vector into actively transcribed *LINE-1* elements. PCR screening was carried out with four established diagnostic primer pairs that were designed previously to bind the consensus ORF1 sequence of *LINE-1* elements (figure 3.47). However, PCR analysis on four clones with consensus primers did not yield expected products for any junctions (data not shown). To determine the genomic location of seamless vector transgenesis in the hESC, genomic DNA from the clones was with digested enzyme *AflIII*, self-circularized (to form a circular DNA) and then subjected to inverse PCR using oppositely oriented primers in the seamless vector as shown in figure 3.48. Unfortunately, no expected products were obtained from the inverse PCR for both left and right junctions in any of the clones.

Figure 3.48. Inverse PCR analysis for identification of the genomic location of seamless vector transgenesis. Schematic diagram showing *AflIII* digestion of recombinant products of seamless vector transgenesis at *attH4X* site of actively transcribed *LINE-1* elements. Positions of relevant primers used for inverse PCR to determine left and right junctions are indicated.

3.4.4. Transgene insertion site mapping to identify the genomic location of *attL4X-GFP-EF-Puro-IRES* transgenesis in hESCs

Since inverse PCR failed to identify the genomic location of seamless vector (*attL4X-GFP-EF-Puro-IRES*) insertion, transgene insertion site mapping was carried out via a novel PCR strategy as shown in figure 3.49 a. Genomic DNA of four clones obtained from co-transfection of 100ng of *attL4X-GFP-EF-Puro-IRES* and 1μg of *pEFssInt-C3CNLS* was used for transgene insertion site mapping and genomic DNA from wild-type hESCs was used as a negative control. Both primary and secondary PCRs on four clones using primers in the seamless vector and adaptor (as shown in step 4 and 5 of figure 3.49 a) did not yield any specific products (figure 3.49 b). Hence, the genomic location of the seamless vector transgenesis could not be determined. Further studies need to be conducted on these clones to identify genomic loci of the seamless vector insertion.

Figure 3.49. Seamless vector genomic insertion site mapping via sequencing. **a)** An illustration depicting the PCR strategy for mapping insertion sites of seamless vector *attL4X-GFP-EF-Puro-IRES* in hESCs. This includes generation of targeted ssDNA using a biotinylated primer complementary to the inserted vector sequence (step 1), capture of the ssDNA with streptavidin beads, ligation of adenylated oligonucleotide (adaptor) and PCR amplification and sequencing analysis (steps 2-6). Positions of the relevant primers used to map *attL4X* (right junction) are indicated in step 5 [adapted with permission (181)]. **b)** PCR analysis for mapping insertion sites from clones obtained from co-transfection of seamless vector *attL4X-GFP-EF-Puro-IRES* with expression plasmid Int-C3. PCR with the primers GFP fwd 599 and petF2RC (see step 5 in figure 3.49 **a**), using genomic DNA from clones obtained co-transfections of *attL4X-GFP-EF-Puro-IRES* and *pEFssInt-C3CNLS* didn't yield any band of the expected size (>930bp). L, 1kb ladder; W: no DNA template control; ES, genomic DNA from parental hESCs; #1, #3, #4, #5: genomic DNA from clones obtained from co-transfection of 100ng of *attL4X-GFP-EF-Puro-IRES* and 1µg of *pEFssInt-C3CNLS*; #5: genomic DNA from clones obtained from co-transfection of 200ng of *attL4X-GFP-EF-Puro-IRES* and 1µg of *pEFssInt-C3CNLS*.

4. DISCUSSION

4.1. λ -Integrase mediated seamless vector transgenesis platform for therapeutic protein expression

One of the main challenges faced by genome engineering tools is to achieve efficient and precise insertion of multi-transgene cassettes at a specific safe locus in the human genome. Efficient and safe transgenesis in the human genome has several applications in the field of gene therapy, regenerative medicine, synthetic biology, functional genomics and biologics production. In that respect, several DNA editing tools, such as plasmids, viral vectors, transposons, engineered nucleases and site-specific recombinases, have been utilized for various applications (43,361). However, engineered nucleases and site-specific recombinases showed substantial advantages over viral vectors with respect to safety, specificity, cost, sensitivity and off-target effects (362). Hence, several studies have been conducted to improve and develop new non-viral tools for transgenesis in the human genome. The engineered nucleases, such as ZFNs and TALENs, have been widely used to introduce different genetic modifications in the genome, that include gene addition, gene disruption and gene corrections. For example, ZFN-induced HDR mechanism has been used to correct disease-causing mutations of hemophilia B, X-linked severe combined immune deficiency (SCID), α 1- antitrypsin deficiency, sickle cell disease, Parkinson's disease and leukemia (56,73,363-366). In addition, ZFN-induced NHEJ-mediated repair has been used to disable the *CCR5* gene in primary T cells, thus conferring HIV-1 resistance as a powerful tool to fight HIV/AIDS (367). Besides ZFN and TALEN, the CRISPR-Cas9 system is the most commonly used tool for human genome editing. The CRISPR-Cas9 system has been broadly utilized in gene and cell therapy for the treatment of diseases such as cystic fibrosis, retinitis pigmentosa, Duchenne muscular dystrophy (DMD), β -thalassaemia, HIV and cancer (368-373). However, issues like insertional mutagenesis, immunogenicity, off-target events, and small payload size have limited its downstream applications (43,374).

Such a novel λ -Integrase mediated transgenesis platform was developed in the laboratory of Professor Dr Peter Dröge, where a λ -integrase (λ -Int) variant can catalyze conservative site-specific integration of a large plasmid into a safe harbor site, termed as *attH4X* (figure 3.1 a). The *attH4X* site comprises 18bp, is present in open reading frame 1 (ORF1) of a small subset of human Long INterspersed Elements-1 (*LINE-1*) dispersed at 933 different genomic loci. It was also found that the *attP4X* sequence on target vector can be functionally replaced by the shorter 120bp *attL4X* hybrid recombination site for Int-mediated recombination with *attH4X* in *LINE-1* (figure 3.1 b). This tool was a new addition in the human DNA engineering toolbox (1).

One of the major requirements of this tool is a circular plasmid (target vector) carrying 241bp of *attP4X* site to site-specifically target genomic *attH4X* via λ -Integrase mediated intermolecular recombination. Thus, the target vector carrying *attP4X* site was required to be produced using

Escherichia coli. This, in turn, would require the inclusion of bacterial backbone along with transgene in the target vector construct for its amplification in *Escherichia coli*. Several studies had shown that the bacterial sequences in the plasmid might lead to transgene-silencing, activation of immune response and also adds on to the size of the plasmid (191,192,234). Thus, it was an absolute necessity to eliminate the bacterial backbone from the target vector to improve the λ -Integrase transgenesis platform in terms of safety and efficiency. With this respect, one of the aims of the study is to generate seamless vector or minicircle (supercoiled plasmid without bacterial sequence) and explore its utility in λ -Integrase mediated transgenesis platform for different applications.

The strategy to generate seamless vector involves an *in vitro* intramolecular recombination on a substrate plasmid carrying direct repeats of *attP4X* and *attH4X*, by a variant of λ -Integrase, Int-h/218, as shown in figure 3.2. Such recombination results in the formation of two interlinked DNA circles with one ring having the payload and *attL4X* site and another ring having the unwanted bacterial backbone and *attR4X* site. The supercoiled seamless vector can be isolated from the interlinked DNA circles by endonucleolytic digestion of the undesired bacterial DNA circle.

The seamless vectors generation via *in vitro* *attH4X* x *attP4X* intramolecular recombination requires purified Int protein. Thus, the C-terminally histidine-tagged Int-h/218 and Int-C3 were overexpressed in *E. coli* (figure 3.3). The cell extracts for both the Int variants were tested for recombination activity by catalyzing *in vitro* intramolecular recombination on substrate plasmid that carries direct repeats of *attP4X* and *attH4X* (figure 3.4 a). It was observed that the recombination efficiency of the cell extracts having Int-h/218 was enhanced to ~45% in the presence of scIHF2 (figure 3.4 b). Even though the Int variants, Int-h/218 and Int-C3, are known to be active in the absence of accessory protein cofactors (178), yet the addition of purified scIHF2 has drastically improved the intramolecular recombination efficiency. Studies have reported the essential role of IHF as a cofactor for phage λ mediated site-specific integrative and excisive recombination (375,376). The enhanced recombination in the presence of IHF is possibly due to its binding to the arm sites of *attP* that induced the DNA bending. This brought the distal ‘arm-type’ binding sites for Int into close proximity with the ‘core-type’ Int sites, and thus enhanced the recombination reaction (167). Between the two Int variants Int-h/218 and Int-C3, the cell extract having Int-h/218 showed higher recombination efficiency of ~75% for *in vitro* recombination (figure 3.5 a, b). Moreover, it was noticed that the activity of the cell extract having Int-C3 slightly reduced over long period storage. Hence, the cell extract having Int-h/218 was selected for further seamless vector generation.

In order to remove other unwanted bacterial proteins from the *E. coli* cell extract that can interfere with *in vitro* intramolecular recombination, C-terminally histidine-tagged Int-h/218 was partially purified using Ni-NTA resin (figure 3.6 a, b, c). A total yield of ~5-6mg of highly active Int-h/218 protein was obtained from 200ml of *E. coli* culture that was safely stored at -20°C. The

recombination activity of purified Int-h/218 protein was analyzed from low to moderate salt concentrations, as it is well known that the DNA-protein interactions are salt sensitive due to electrostatic interactions (267). Surprisingly, upon testing the purified Int-h/218 protein for recombination activity, it was noticed that high salt concentration (~150mM KCl) in the reaction buffer has improved the recombination efficiency (figure 3.7 b). The enhanced activity of Int-h/218 in high salt concentration also correlates with the seamless vector generation employing crude cell extract, where the reaction mixture also contained 150mM KCl upon using 10 μ l of cell extract. It was found that only a low concentration of protein (15ng/ μ l) was required to obtain ~40-50% of recombination efficiency to generate seamless vector in the presence of scIHF2. Hence, a simple, highly efficient protocol was developed and optimized for the generation of a seamless vector with partially purified Int-h/218 having ~40-50% recombination efficiency in the presence of 150mM KCl. This could provide enough amount of purified seamless vector required for *ex vivo* targeting of the human genome with seamless vector.

In order to further study the *in vitro* recombination by Int-h/218, intramolecular recombination was carried out on a linearized substrate plasmid that yielded covalently closed relaxed seamless vector in contrast to the negatively supercoiled seamless vector obtained from a non-linearized supercoiled plasmid (figure 3.8 b). Interestingly, it was found that the recombination efficiency of Int-h/218 mediated recombination on the linearized plasmid was higher than the efficiency obtained by supercoiled substrates. Studies have shown that wild-type integrase requires DNA supercoiling of target sites to catalyze recombination (167,377). However, the mutant Int variant, Int-h/218 can perform both integrative and excisive recombination in the absence of negative DNA supercoiling *in vitro* (173,178,378). This might be one of the reasons for efficient recombination by Int-h/218 on a linearized plasmid in the absence of any DNA supercoiling.

In addition, it was found that Int-h/218 mediated *in vitro* intramolecular wild-type *attP* and *attB* recombination is preferred over intramolecular *attL* and *attR* recombination (figure 3.9 c). A similar finding was observed by Tan and Dröge (2005), where Int-h/218 catalyzed the integrative pathway by *attP* and *attB* recombination, five times more efficiently than the excisive pathway by *attL* and *attR* recombination for site-specific recombination in human embryonic stem cells (379). This suggests that the precise arrangement of Int arm binding sites in *att* sites is important for recombination (1) and the efficiency of excision reaction by factor-independent Int variants is low in the absence of XIS.

To expand the λ -mediated seamless vector production to a large scale, an *in vivo* platform was employed to scale up the generation of the seamless vector. Such an approach will make the seamless vector production platform more robust and efficient for its utility in the biopharmaceutical sector. Even though the optimized *in vitro* protocol for the generation of the seamless vector was good

enough in terms of DNA yield for routine *ex vivo* applications, however, this protocol was not suitable for large scale seamless vector production. The *in vitro* procedure for seamless vector generation is laborious, time-consuming and requires continuous purification of Int-h/218 protein in every 6 months as the protein activity diminishes over long periods of storage. Hence, to overcome these issues, the platform for seamless vector production was shifted to an *in vivo* system using *E. coli* strains LZ41 (figure 3.10) (269,272-274). This strain expresses wild type λ -integrase which is tightly controlled by a thermolabile repressor (cI857). Thus, the integrase expression could only be achieved at 42°C, however, due to thermosensitivity, the Int mediated recombination is not active at 42°C. Kotewicz *et al.* (380) had shown that the wild-type Int protein exhibited temperature sensitivity for DNA binding from 35°C to 42°C. Thus, the temperature of the *E. coli* culture needs to be returned at 30°C to obtain recombination activity with wild-type λ -integrase (269).

Both the strains LZ41 and LZ54 could successfully generate seamless vector from the substrate plasmid carrying direct repeats of *attH4X* and *attP4X* (figure 3.11 a). Even though the *attP4X* and *attH4X* had a different 7bp overlap sequence as compared to the wild-type *attP* and *attB* sequence (1), yet the wild-type λ -integrase can recognize these sites and successfully perform intramolecular recombination on this substrate with high efficiency. The overlap sequence of *attP4X* was modified in the previous study to match the overlap sequence of *attH4X* in the human genome (1).

To generate a seamless vector, both the strains were transformed with a substrate plasmid and the recombinant products were digested with a restriction enzyme and analyzed by agarose gel electrophoresis (figure 3.11 b). In the induced culture, a mixture of unrecombined linear substrate plasmid, linear bacterial backbone, and supercoiled seamless vector was observed (figure 3.11 b). Besides the expected recombinant products, there was a prominent linearized product of about 11kb (figure 3.11 b). Upon analyzing the linear product by appropriate restriction enzyme digestion (figure 3.11 c), it was found that it corresponds to the linear intermolecular recombination product of *attL4X* on seamless vector and *attH4X* on the substrate plasmid (figure 3.11 c). It further suggests that wild-type integrase could also recognize *attL4X* hybrid recombination site as recombination partner for *attH4X* site on substrate plasmid. Moreover, the wild-type integrase could perform equally efficient recombination in both *cis* (intramolecular recombination on substrate plasmid) and *trans* (intermolecular recombination between seamless vector and substrate plasmid). Even though the chance of the interaction of wild-type integrase protein with *att* sites on substrate plasmid is much higher in *cis*, yet efficient intermolecular recombination in *trans* suggests that intracellular environment might have positively influenced the intermolecular recombination. However, Int mediated intermolecular recombination between seamless vector and substrate plasmid is disadvantageous for the system as it will reduce the total yield of the seamless vector. This loss of

seamless vector cannot be recovered; as the seamless vector generation will increase, it will simultaneously drive the intermolecular recombination between seamless vector and substrate plasmid.

Overall, both the strains LZ41 and LZ54 could successfully perform *in vivo* intramolecular recombination on substrate plasmid to produce the seamless vector. However, it was found that the integrase mediated recombination efficiency in strain LZ41 was higher (~60%) than the efficiency obtained in strain LZ54 (figure 3.11 b). Nevertheless, both the strains genetically vary with each other that might have contributed to the difference in integrase mediated recombination efficiency. Strain LZ41 carries *parC*^{K84} mutation in *parC* gene and strain LZ54 carries *gyrA*^{L83} mutation in *gyrA* gene that makes the topoisomerase IV subunits and DNA gyrase respectively highly resistant to norfloxacin drug (269,272-274). This drug is known to block DNA decatenation in wild-type *E. coli* strains. Since norfloxacin drug has been not added to the *E. coli* culture, thus, the interlinked DNA circles formed by intramolecular recombination have been decatenated by topoisomerase IV, resulting in the formation of two separate supercoiled molecules of seamless vector and bacterial backbone as shown in figure 3.11 a.

The cell culture condition of LZ41 was further optimized for efficient *in vivo* seamless vector production. During optimization of protocol for seamless vector generation in strain LZ41, it was observed that a high *in vivo* recombination efficiency of ~85% was achieved upon incubating the culture at room temperature as compared to 30°C after the induction process (figure 3.12, 3.13). After testing different combinations of cell culture conditions, the protocol for the generation of a seamless vector using strain LZ41 was optimized at 42°C for 30mins for induction, followed by incubation at room temperature for 3hours without shaking. This optimized protocol was further tested with two substrate plasmids of different sizes that achieved a recombination efficiency of ~50% for seamless vector generation (figure 3.13). As a proof of concept, *in vivo* seamless vector production platform was also utilized to generate seamless vector encoding a therapeutic protein, T-cell Receptor (TCR), to demonstrate its application in the biopharmaceutical sector. Upon scaling-up, the volume of culture to 100ml, a recombination efficiency of ~50% was achieved which yielded 20µg of the seamless vector (figure 3.14). It further demonstrates the utility of the platform for large scale production of seamless vector for biopharmaceutical industries.

The seamless vector carrying *attL4X* was validated for site-specific integration at the genomic *attH4X* site within *LINE-1* elements in HT1080 cells and human embryonic stem cells (hESCs) (figure 3.15, table 5, 9). The *attL4X-PGKssPuro* seamless vector was first targeted in HT1080 cells and insertion of the seamless vector into *LINE-1* elements were confirmed by PCR analysis on the HT1080 bulk cultures (figure 3.16, table 6). Although, each bulk culture contained cells from more than 100 colonies, yet specific genomic loci for left and right junctions were identified by PCR

analysis (table 6). It might be due to biasness in PCR amplification using four primer pairs that were designed to bind the consensus ORF1 sequence of *LINE-1* elements. Moreover, different genomic loci for left and right junctions were identified by sequencing analysis (table 6) that indicated multiple *LINE-1 elements* targeting with seamless vector in each bulk culture.

In order to determine different *LINE-1* loci of seamless vector transgenesis in the HT1080 bulk culture, a novel PCR strategy was utilized (figure 3.18 a). Deep paired-end sequencing on the obtained PCR product (figure 3.18 b) yielded 434 different *LINE-1* loci where seamless vector integration had occurred in bulk cells (Appendix A1). Out of 434 *LINE-1* loci, a genomic coordinate at chromosome 10, located at intergenic region represented the highest count of 14116 that could suggest frequent targeting of the seamless vector at this *attH4X* locus. Besides, 3/4th of these loci was counted only thrice or less, which could indicate that these loci were less frequently targeted via Int-C3 mediated seamless vector recombination. Thus, this strategy allowed the identification of 434 different *LINE-1* loci out of 933 *attH4X*-containing *LINE-1* that had been targeted at least once by the seamless vector.

Additionally, the clonal population obtained by the *attL4X-PGKssPuro* seamless vector transgenesis in HT1080 cells and human embryonic stem cells (hESCs) were also analyzed. PCR screening showed a targeting efficiency of 50% for HT1080 cells (figure 3.17 b, c, d) whereas the efficiency in hESCs reached as high as 78% without generation of any clones for Inactive Int (figure 3.21 b, c, d, e). PCR and sequencing analysis of individual clones of HT1080 and hESC allowed to identify genomic loci of site-specific insertion of the seamless vector in the human *LINE-1* element (table 7, 10). Almost half of the identified genomic loci correspond to intergenic regions while the other half belonged to the introns of genes. Since about 1/3 of the 933 *attH4X*-containing *LINE-1* are present in the intron of genes (181), hence, it suggests that the site-specific insertion by λ Int is influenced by the local chromatin structure and epigenetic status of the genome. In addition, a significant fraction of HT1080 clones had shown different genomic *LINE-1* loci for left and right junctions of integration indicating multiple copy insertion of the seamless vector into the genome (table 7, 10). However, in six HT1080 clones, #27, #41, #45, #46, #46, #80, same genomic loci were identified from left and right junctions indicating a single copy of seamless vector integration at *LINE-1* element (table 7). Similarly, for hES clones, different genomic *LINE-1* loci were identified for left and right junctions that might suggest multiple copy insertion of the seamless vector at *LINE-1* element (table 10). Interestingly, a specific locus at chromosome 2 was identified in five hES clones for left junction (table 10) suggesting biasness in PCR amplification using consensus primer cs_attH4X_F2.

The use of seamless vector for transgenesis in HT1080 and hESCs had improved the efficiency of site-specific targeting and reduced the random transgene insertion events. This could be

exemplified using the targeting data obtained with HT1080, where 45/92 clones were positive for *attH4X* targeting with *attL4X-PGKPuro* seamless vector (table 7) compared to 4/33 clones with the previous study of *pattP4X-PGK-Puro* plasmid-based targeting (1). Additionally, the seamless vector transgenesis resulted in only 2 clones with the inactive version of the integrase variant suggesting lower rates of off-target effects. Similarly, for hESCs, no clone was obtained by co-transfection of the *attL4X-PGKPuro* seamless vector with Inactive Integrase variant suggesting minimal off-target effect with seamless vector transgenesis. Moreover, the seamless vector was site-specifically targeted in 25/32 hES clones (table 10) as compared to a previous study where site-specific targeting of an 8kb plasmid at genomic *attH4X* was positive only in 3/68 hES clones (1). Thus, switch from plasmid-based targeting to seamless vector transgenesis improved the efficiency of the platform for site-specific integration of seamless vector in the *LINE-1*. The elimination of unnecessary bacterial sequences from plasmid provides an immense opportunity to increase the size of transgene cassette for gene delivery in human cells making it suitable vast range of applications. In addition, it can be utilized for insertion of large transgene along with complex genetic regulatory control regions, such as silencers and enhancers, site-specifically into the genome.

Different possible recombination scenarios for co-transfection of substrate plasmid with Int-C3 were explored in HT1080 cells (figure 3.19). Transfection of substrate plasmid with Int-C3 in HT1080 cells resulted *in vivo* intramolecular recombination between *attP4X* and *attH4X* on substrate plasmid leading to seamless vector generation. It was followed by insertion of the seamless vector in the genome via *attL4X* and genomic *attH4X* intermolecular recombination (figure 3.19 right panel). The PCR and sequencing analysis revealed that 43% of the expanded HT1080 clones had seamless vector integration into human *LINE-1* element (figure 3.20 a, b) while only 20% of the clones had substrate plasmid integration (figure 3.20 c, d). This indicated preferability of *in vivo* λ Int-mediated intramolecular recombination between *attP4X* and *attH4X* on substrate plasmid over intermolecular recombination between *attP4X* and genomic *attH4X*. This is possibly due to the high probability of random collision between *attP4X* and *attH4X* on the substrate plasmid compared to *attP4X* on substrate plasmid and distant *attH4X* in the genome (396). In other words, it might lead to increase in productive encounters between *attP4X* and *attH4X* on the substrate plasmid resulting in the formation of seamless vectors and hence a better chance of targeting of the human genome with seamless vectors

To explore the application of the seamless vector for therapeutic protein expression, the λ -integrase mediated seamless vector transgenesis platform was utilized to achieve insertion and expression of anti-CD19 chimeric antigen receptor (CAR) transgene in human embryonic stem cells (hESCs) as proof of principle (figure 3.23 a). The anti-CD19 CAR transgene is mostly expressed on tumour-reactive T cells to develop a potent and powerful treatment for B-cell leukemia and

lymphoma (figure 3.22 b). However, the high cost of generation of autologous CAR-T cell has limited its usage economically impractical (293). Thus, integrase -mediated transgenesis of anti-CD19 CAR cassette in hESCs can act as a basis for CAR-T-based immunotherapy where successfully targeted hESCs can be differentiated into CAR-T-cell lineage for allogenic immunotherapy of B-cell leukemia and lymphoma (295,381,382). For this purpose, seamless vector *attL4X-PGKssPuro-EF-CAR* was delivered in the hESCs where two hES clones were identified that showed seamless vector insertion in the human *LINE-1* element (figure 3.23 a, b, c). Surprisingly, even though there was a complete integration of the seamless vector in the genome for both the clones, they lacked anti-CD19 CAR expression. It could be due to the integration of seamless vector at epigenetically silenced *LINE-1* elements in hESCs. It is known that *LINE-1 elements* in germ cells (383) and endogenous retroelements (384) are silenced due to the presence of repressive histone modifications (385). Hence, λ Int-mediated transgene insertion into any of this silenced *LINE-1* in hESCs can lead to partial or complete silencing of the transgene expression.

In a parallel approach, another seamless vector *attL4X-PGKss-CAR-EF-Puro* was employed for transfection in hESCs (figure 3.24 a). In this case, the promoter to express anti-CD19 CAR transgene was switched from EF α to PGK promoter to obtain enhanced and stable transgene expression in undifferentiated hESCs (298). A total of 12 hES clones out of 48 clones were obtained with this approach that had seamless vector insertion in the human *LINE-1* element were obtained (figure 3.24 b, c, table 11). However, only two clones showed anti-CD19 CAR expression in FACS when stained with anti-human CD3 ζ primary antibody. FACS analysis showed 87.1% and 73.3% of CAR-expressing cells in clone #19 and #34 as compared to non-transfecting cells (figure 3.24 e). Since clone #34 could not survive in the cell culture conditions, clone #19 was further characterized by western blot which confirmed anti-CD19 CAR expression from the stable seamless vector integration in the genome (figure 3.25). In addition, the Southern Blot analysis of clone #19 revealed a single copy of the transgene integration that lead to such a high expression of anti-CD19 CAR (figure 3.26 b, c). Moreover, clone #19 showed stable CAR transgene expression over a period of 6-8 weeks (figure 3.27) indicating the potential utility of seamless vector for insertion and expression of the therapeutic protein in the field of gene therapy.

Currently, different strategies, such as the CRISPR-Cas9 system, transposons, meganucleases, are being used to engineer T-cells for anti-CD19 CAR expression. For example, *piggyBac* and *Sleeping Beauty* were utilized to randomly insert CAR transgene in the T-cell genome via electroporation (386,387). Although, the strategy had shown a lower level of T-cell toxicity, yet the requirement of a large amount of DNA for transfection and activation of oncogene due to insertional mutagenesis had affected its applications in CAR-T cells generation (387). In addition, homing endonucleases, such as ZFNs and megaTAL (fusion of meganuclease with the DNA binding region

of TALEN), were also used to site-specifically integrate CAR transgene into the *CCR5* locus via HDR (388). Site-specific insertion of the CAR transgene at *TRAC* locus of T-cell genome was achieved by CRISPR-Cas9 induced HDR that resulted in uniform expression of CAR transgene from the endogenous regulatory elements (389,390). However, several issues like off-target effects, activation of p53-mediated DNA damage response, insertional mutagenesis, generations of indels have questioned the safety of CRISPR-Cas9 edited CAR-T cells (391-393). A recent study had reported the off-target effects generated by insertion of CAR transgene at *TRAC* locus by CRISPR-Cas9 mediated HDR that has led to the activation of oncogenes (394). Besides another study also showed that the CRISPR-Cas9 mediated genome editing has resulted in hundreds of unintended mutations throughout the genome of a mouse model (395).

In this respect, the λ -integrase mediated seamless vector transgenesis tool can act as an alternative to CAR-T based immunotherapy where λ -integrase mediated seamless vector transgenesis could be used to first engineer autologous, self-renewable T-cell-derived iPSCs with multiple CARs transgene, followed by *ex vivo* differentiation of the engineer cells to CAR-T cell lineage for treatment of B-cell leukemia and lymphoma. As a proof of concept, this platform was utilized to target two-reporter cassette in hESCs at *LINE-1 elements* and achieved sustained expression of therapeutic protein in ~90% cells in a clonal population. It provides an example of the usefulness of the tool to integrate and express multi-transgene cassette to overcome various clinical problems in the therapeutics fields. This tool provides benefits of high targeting efficiency along with highly sustained transgene expression and minimal off-target effects, thus demonstrating the suitability of this platform for therapeutic application. In addition, it can be further expanded to *in vivo* gene therapy using nano-carriers as transgene vehicle.

Five independent transfections in HT1080 cells and hESCs has allowed the identification of several *LINE-1* that have been targeted via Int-C3 mediated *attL4X* recombination at genomic *attH4X*. Thus, upon analyzing data from five lists of targeted *attH4X* sequences obtained in HT1080 cells and hESCs (tables 7, 8, 10, 11, Appendix A1), 27 loci, also termed as ‘hot-spot’, were identified as preferred *LINE-1* loci for Int-C3 mediated seamless vector transgenesis (table 13). The ‘hot-spot’ was targeted at least twice in four independent cell transfections (table 13) and identified separately for each cell line. It was observed that almost half of the identified ‘hotspots’ belongs to introns while the rest are in intergenic regions. This further indicates that the target site selectivity for Int-C3 mediated recombination might be dependent on the local chromatin status. Considering that only one-third of the 933 *attH4X*-containing *LINE-1* is actually located in introns (181), it is most likely that the transcriptional or epigenetic status of *LINE-1* plays a vital role for the recombinogenic potential of Int-C3 with the human genome (1). However, the *LINE-1* that has been targeted in all independent transfections in HT1080 cells and hESCs is in intergenic regions on chromosome 3. This locus might

act as a safe harbor site for Int-C3 mediated transgenesis. The DNA sequence logo on these 27 ‘hot-spot’ showed adenosine preference in the first and third position and cytosine preference in the second position of the first three variable nucleotides at 5’ end of 21bp *attH4X* sequence (below table 13). Such specificity in the target site sequence might positively influence Int-C3 mediated *attL4X* recombination with the genomic *attH4X* site. Thus, in future, the Int variants could be further engineered to enhance its recognition for this subset of *attH4X* sequence (179,180).

The current study is the first example of λ Int-mediated human genome editing platform that manifests dual capability to generate seamless vectors as well as target the human genome for insertion and expression of transgene cassette. This platform is simple, efficient and highly productive in terms of generation of seamless vectors for *ex vivo* applications. It also allows insertion and sustained expression of large transgene cassette in the human genome by the elimination of bacterial backbone from the target vector, which is known to silence transgene expression. This platform has a high potential for applications in gene therapy and biologics production.

4.2. λ -Integrase mediated seamless vector transgenesis platform for biologics production

λ -integrase mediated transgenesis tool was further utilized in HEK293T cells for biologics production. Human embryonic kidney 293 (HEK-293) cells are the most popular mammalian expression system for recombinant protein production. As a protein expression system, HEK293 cells provides several benefits including easy efficient transfection of the transgene, human post-translation modifications to the protein, and high expression of the transgene. The strategy designed for biologics production platform in HEK293T cells involves the generation of a “master cell line” by randomly integrating a landing pad carrying four copies of *attH4X* and a hygromycin reporter cassette, followed by λ -integrase mediated targeting of the payload at *attH4X* sites of the landing pads (figure 3.28 a). The four copies of *attH4X* were included to increase the local concentration of *attH4X* recombination site at a single locus for an enhanced recombination efficiency with Int-C3.

Local concentration is considered as a basic principle of life (397). It is a phenomenon, involved in all kinds of chemical reaction that also includes specific DNA-protein, protein-RNA and protein-protein interactions. Such interactions between molecules could be competitive or cooperative in nature. The rate of chemical reactions is highly dependent on the local concentration of the molecules involved (398). Dröge and Müller-Hill (2001) proposed that the regulation of local concentration of protein molecules inside the cells is one of the main mechanisms behind the control of gene expression and other DNA interactions (398). The phenomenon of local concentration was further classified into two types based on the mode of action (398). In local concentration type I reaction, a protein bound to a specific DNA sequence increases local concentration for another protein-bound DNA molecule or a secondary DNA molecule in close proximity. It could be further

explained by binding of transcriptional activator/repressor proteins at the specific DNA sequence located within 1kb upstream of promoter sequence. This leads to increases the local concentration of the regulatory protein in the vicinity of the promoter, thus regulating the gene expression in eukaryotic cells (398). The type II local concentration reaction involves the formation of nucleoprotein complexes that would otherwise not form in intracellular solution. Such complexes establish long-range interaction with similar complexes located at a distal position in the eukaryotic genome (398). For example, binding of transcription factors on the enhancer elements located distally (1-100kb upstream/downstream of the gene) to form a nucleoprotein complex. Such complex would not have existed in intracellular solution in the absence of local concentration effect. This nucleoprotein complex further positively interacts with another DNA element (promoter sequence) in *cis* or *trans* to trigger transcription in eukaryotic cells (398). Besides the above-mentioned examples, Müller-Hill, 2006, also proposed that the hydrogen bond interaction between two bases in a DNA molecule and catalyzation of chemical reactions by enzymes are some examples of the effect of increase in local concentration of the interacting molecules (397,399).

In this respect, the four copies of *attH4X* were included in the landing pad to increase the local concentration of the nucleoprotein complex formed by integrase protein and *attH4X* sequence to attract more molecules of integrase bound seamless vector for site-specific recombination at the landing pad. Moreover, the DNA sequence of *attH4X* in the landing pad is derived from the DNA sequence logo obtained from 27 ‘hotspots that are identified as preferred *LINE-1* loci for seamless vector transgenesis. This, in turn, will positively influence Int-C3 mediated *attL4X* recombination with the *attH4X* site of the landing pad.

Random insertion of landing pad resulted in 93 docking clones that harboured at least one copy of the landing pad. Upon targeting the landing pads with seamless vector (*attL4X-Puro-EF-GFP*) and Int-C3 (figure 3.28 b), 18 targeted clones were analyzed by PCR based on their high EGFP expression. The PCR analysis showed that 9/18 clones harboured at least one copy of the seamless target vector targeted at the landing pad (figure 3.29 b, c, d). Interestingly, even though HEK293T genome carries 933 endogenous *attH4X* sites for Int-C3 mediated recombination (1), yet the seamless vector was site-specifically targeted at the artificially introduced *attH4X* sites of the landing pad suggesting that the local concentration of the recombination site at a single locus influenced integrase mediated recombination at the landing pads.

The FACS analysis of three clones (#3T, #11T, and #18T) revealed that ~90% cells of the clones had a high homogenous expression of EGFP (figure 3.30). In addition, Southern blot analysis on the parental HEK docking clones revealed that the linear landing pad formed concatemers which were integrated into a single genomic locus for clone #3 and #18 and two genomic loci for clone #11 (figure 3.31 a). This suggested that the parental HEK docking clones harboured at least two copies of

the landing pads and the integrase mediated seamless vector integration was biased towards artificially introduced ~10 copies of *attH4X* sites in the landing as compared to 933 endogenous *attH4X* sites in the genome. Due to time constrain, only one targeted clone, #3T, was further analyzed with a Southern blot, which revealed that only one copy of the seamless target vector was integrated into the landing pad (figure 3.31 b). This further suggests that the landing pad for the clone might be present in the open chromatin region which had led to high expression of EGFP from landing pad even with a single copy of seamless vector integration. Absence of any signal other than expected 5.1kb band on the southern blot using probes complementary to *Puro* and *EGFP* gene (figure 3.31 b), indicated no other recombination event that had occurred at the endogenous *attH4X* sites of the genome. This further supports the local concentration effect of *attH4X* sites for integrase mediated recombination at the landing pads. Finally, a master cell line, HEK Doc #3 was selected for recombinant protein production.

The biologics production platform in HEK293T was further utilized to integrate full-length factor 8 (FLF8) gene and express factor VIII protein. Factor VIII (FVIII) is a vital blood-clotting protein which is in high demand for its production in the biopharmaceuticals sector as an approach to treat hemophilia A patients. Current biologics production platforms, such as viral systems and transposons, fail to efficiently deliver and insert full-length factor 8 (FLF8) gene of size 8kb, in the mammalian cells. In this respect, λ -integrase mediated biologics production platform can be advantageous over other platforms due to the inherent ability of wild-type λ -integrase to integrate its complete genome of size 48kb site-specifically at the bacterial genome with high efficiency. Thus, as a proof of concept the biologics production platform was utilized for factor VIII protein production to test its utility in the biopharmaceuticals production sector.

The master cell line, HEK Doc#3, was used for insertion and expression of the FLF8 vector, *pattP4X-Puro-EF-FLF8-IRES-ZsGreen-attH4X* (FLF8 vector), of size 14.2kb that encodes therapeutic factor VIII protein and other reporter cassettes (figure 3.32). The transfection of FLF8 vector with Int-C3 in HEK Doc #3 resulted in the complete insertion of the 14.2kb vector at the landing pad which was further confirmed by PCR analysis (figure 3.33 c, d, e). The HEK293T bulk culture obtained from this transfection was analyzed by FACS and found that only 5.6% of cells were ZsGreen+ (figure 3.34 a, left panel). Hence, these populations were enriched to 85% by FACS sorting (figure 3.34 a, right panel). This enrichment was further confirmed by genomic PCR for transgene insertion and tested for expression of the factor VIII protein (figure 3.34 b).

The factor VIII activity test of the sorted bulk culture was measured by a chromogenic assay and the data obtained from the assay were normalized to cell viability (obtained from MTT assay) to acquire an absolute fold change in protein activity over the parental clone. A corresponding value of absorbance for each sample was obtained from two time points (T1 and T2) in the linear range of the

plot. The absorbance value of DMEM media was subtracted from the corresponding absorbance value obtained from each sample. The analysis revealed that the fold change in FVIII protein activity from the unsorted bulk culture was only ~3.5 times higher than the parental HEK docking clone #3 (figure 3.34 b). Interestingly, the sorted bulk culture showed a fold change of ~20 times for factor VIII protein activity over parental clone (figure 3.34 b). Upon plotting the corresponding absorbance value for sorted bulk culture on the standard curve of the FVIIIa Enzyme (figure 3.34 c), the total yield of FVIII protein from the sorted bulk culture was estimated to be ~2.27µg/ml (table 15). This further suggested the enrichment of factor VIII producing cells by FACS as the unsorted bulk culture showed 75-fold lower yield of factor VIII protein. A similar yield of full-length factor VIII was also obtained with lentiviral (400), retroviral (401) and transposon (323) driven expression system in BHK-M cells and HEK293T cells.

Taken together, as a proof of concept, the λ -integrase mediated biologics platform was utilized to insert FLF8 vector of size 14kb that resulted in high expression of functionally active factor VIII protein. It further demonstrates the utility of this platform for the production of recombinant proteins and monoclonal antibodies in the biopharmaceutical sector. Moreover, this platform is suitable for multiple insertions of the transgene at the landing pads of HEK293T cells in order to enhance the yield of recombinant proteins. Since the landing pad in HEK293T cells is positioned at a specific genomic locus, the insertion and expression of a transgene from the landing pad will be devoid of any chromatin position effect. Thus, this platform can also allow testing of different variants of a recombinant protein in order to select the best performing variant. The current study is the first example of λ Int-mediated transgenesis of a large vector of size 14kb in the human genome. Stable integration and expression of a large transgene in the human genome are considered as one of the main hurdles in gene therapy. In this respect, of λ Int-mediated seamless vector transgenesis platform can be further utilized in gene therapy for diseases, such as Duchenne muscular dystrophy (DMD) and cystic fibrosis, that require insertion and expression of a large transgene in the human genome. Although several attempts were taken with both plasmid and seamless vector to obtain a clonal population for HEK docking clones targeted with FLF8 vector, none of them was successful (table 16,17). Thus, further optimization is required in the FLF8 targeting strategy. In this respect, the FLF8 vector can be co-transfected with Int-C3 and scIHF2 expression plasmids for site-specific integration of FLF8 vector in the landing pad. As mentioned above IHF plays an essential role as cofactor for phage λ mediated site-specific integrative and excisive recombination (375,376). Therefore, co-transfection of IHF expression plasmid along with Int-C3 might boost the integrative recombination between *attL4X* on FLF8 vector and *attH4X* on the landing pad. In addition, another Int variant, Int-h/218, could also be used along with IHF expression plasmid for co-transfection with FLF8 vector in HEK docking clones. The integrative recombination between *attL4X* and *attH4X* might enhance in

the presence of Int-h/218 and IHF. With these two strategies, it might be possible to obtain a clonal population of HEK docking cells targeted with FLF8 vector.

4.3. Enhancement of transgene expression from seamless vector carrying CTCF binding motifs

Besides exploring different applications of λ -integrase transgenesis platform, it was also necessary to make the platform more efficient in term of transgene expression. It is well studied that the transgene expression is dependent on the chromatin status of the site of the integration (230). Thus, several studies were conducted to find a solution to overcome the chromatin position effect and to obtain sustained and high transgene expression irrespective of chromatin status. Various regulatory DNA elements such as locus control regions (LCR), ubiquitous chromatin region opening element (UCOE), insulators and scaffold/matrix attachment regions (S/MARs) have been studied so far to diminish chromatin position effect on transgene expression (234).

Another DNA element that has not been well studied with respect to its effect on transgene expression is CCCTC-binding factor (CTCF). It is a transcription factor that can perform multiple functions in the cell, such as, regulation of gene expression by promoter-enhancer interaction; remodelling of the 3D structure of the genome by TADs formation (figure 3.36); alteration of nucleosome positioning, histone modifications and DNA methylation (325,402,403). Studies have shown that CTCF is associated with the formation of a DNA loop structure with the cohesion molecules (figure 3.37). Such topological structure is known to promote the transcription process and increase the promoter-enhancer interaction (334,335).

Taken together all these properties of CTCF, the hypothesize was to study the effect of CTCF binding motif on integrated transgene expression in the genome. To test the hypothesis, 14bp CTCF binding motifs were introduced at different locations of the substrate plasmid. The 14bp DNA sequence was obtained from several ChIP-seq experiments performed to identify CTCF binding motifs in the human genome (325) as shown in figure 3.38 b. Previous studies suggested that the orientation of the CTCF binding sites is crucial for DNA loop formation and a convergent orientation of CTCF motifs is a prominent feature of CTCF mediated DNA loop formation (325,404,405). Therefore, the 14bp DNA motifs were placed in convergent orientation in the substrate plasmid. The motifs were placed in two separate positions of the substrate plasmid: (i) CTCF motifs flanking the transgene cassette *UTF1-mNeon* of size ~2.2kb (figure 3.41 b) (8) CTCF motifs flanking the transgene cassette *UTF1-mNeon-PGKss-Puro* of size ~4kb (figure 3.41 c). Such a design was chosen in order to understand whether the CTCF mediated DNA loop formation is dependent on the length of the DNA vector. Besides these vectors, a negative control vector was designed that carried scrambled CTCF binding site flanking the transgene cassette *UTF1-GFP* (figure 3.41 a).

Upon successful transfection of three seamless vectors in human embryonic stem cells (hESCs), it was noticed that the number of mNeon⁺ clones and their respective percentage of mNeon⁺ cells were

higher for CTCF vectors as compared to control vector as shown in (figure 3.42, table 18). This finding was interesting as it correlated with the findings of Zimmerman *et al.*, where the transient expression of the transgene was enhanced to 14-fold when CTCF binding motif was included in the transgene construct (406). However, PCR analysis using four diagnostic primer pairs that were designed previously to bind the consensus ORF1 sequence of *LINE-1* elements (figure 3.43) showed that none of the clones obtained from three seamless vectors harboured site-specific integration of seamless vector at *LINE-1* elements. Inverse PCR analysis on these clones revealed the genomic location of the seamless vector insertion (figure 3.44 a, b) and confirmed that none of the obtained genomic locations belongs to *LINE-1* elements (table 19).

The findings indicated that the inclusion of the CTCF binding site somehow had affected the λ -integrase mediated recombination between *attL4X* on the seamless vector with genomic *attH4X*. CTCF might have interrupted the random collision of Int bound *attL4X* on the seamless vector with genomic *attH4X*, thus inhibiting the formation of synaptic complex essential for λ -integrase mediated recombination. Another possible reason could be that the presence of CTCF binding site near to *attL4X* on seamless vector did not allow integrase binding on the *P'* arm of the *attL4X*, thus inhibiting the whole recombination process. Further studies need to be conducted on these clones in order to understand how CTCF binding site affected the integrase mediated seamless vector transgenesis. Similar findings were obtained upon a second round of transfection in hESCs with seamless vectors carrying CTCF sites. Unlike last time, there was not much difference in the number of mNeon expressing clones between negative control and vectors carrying CTCF sequences (table 20) suggesting the CTCF binding on the seamless vector might be affected due to chromatin or epigenetic status of the genomic loci of seamless vector integration. Therefore, in the future study, a specific genomic locus should be selected to understand the effect of CTCF binding sites on transgene expression. This could be achieved by utilizing the λ -integrase based biologics platform in HEK293T cells where transgene along with CTCF binding site could be inserted at the landing pad located in a specific genomic to analyze its effect on transgene expression.

4.4. λ -Integrase mediated seamless vector transgenesis in actively transcribed *LINE-1* elements

With an aim to improve the λ -Integrase mediated seamless vector transgenesis tool more specific in term of targeting and transcribing transgene using an internal promoter, the possibility of targeting *attH4X* site in an actively transcribed *LINE-1* element was explored in human embryonic stem cells (hESCs). The *LINE-1* elements make up to ~21% of the human genome which is composed of 5'UTR, ORF1, ORF2, and 3'UTR region (figure 3.45). The untranslated 5'UTR harbors an internal RNA Polymerase II transcription promoter in the sense orientation (343). Several studies have shown a majority of *LINE-1* elements are incapable of retrotransposition due to mutations, rearrangements or

truncations at the 5' end, making them transcriptionally inactive. However, only a small subset of (80-100) of full-length *LINE-1* elements are reported to be functionally active and *LINE-1* RNA had been detected in few cell lines, such as HeLa, NTera2D1, HL60, hESCs and HEK293. This suggests that some of the *LINE-1* elements might be actively transcribed (343,359).

In addition, Chandra *et al.* recently showed the *attH4X* sequence (target site of λ -Integrase) is found in ORF1 of *LINE-1* elements at 933 different loci (1). Hence, the strategy to achieve seamless vector targeting at actively transcribed *LINE-1* element involved the design of the seamless vector (*attL4X-GFP-EF-Puro-IRES*) in such a way that upon λ -Integrase mediated integration at *attH4X* site, it will align in the same orientation of the 5'UTR promoter of *L1* element leading to expression of the promoterless puromycin resistant gene from the endogenous promoter of the *L1* element (figure 3.47). The promoterless puromycin cassette in the seamless vector is preceded with an IRES sequence to make sure that any transcripts that would have been generated upstream of puromycin resistant cassette, does not interfere with the translation and expression of puromycin resistant gene (figure 3.47).

Interestingly, transfection of the seamless vector and Int-C3 in hESCs resulted in only a handful of clones which were less in number as compared to the previous study (section 3.1). This suggested that the approach was being specific and provides fewer non-specific clones. Surprisingly, in genomic PCR analysis, none of the clones was detected for seamless vector insertion in any of the 933 *LINE-1* elements using four established diagnostic primer pairs that were designed previously to bind the consensus ORF1 sequence of *LINE-1* elements (figure 3.47). This might be due to that fact that none of the 933 *LINE-1* elements carrying *attH4X* is actively transcribed from an endogenous promoter and so the diagnostic primer pairs could not detect the genomic location of seamless vector insertion. The seamless vector might have integrated to any other *LINE-1* elements or could have been promoter trapped in a different genomic locus which had led to promoterless seamless vector expression from an endogenous promoter. To determine the genomic location of seamless vector insertion, both inverse PCR (figure 3.48) and transgene insertion mapping (figure 3.49 a, b) were performed with the genomic DNA of the puromycin resistant clones and wildtype hESCs as a negative control. However, none of these efforts was able to detect the genomic location of the seamless vector insertion which might be due to repetitive and variable nature of *LINE-1* DNA sequences (407) which made it difficult for PCR amplifications. Further studies need to be carried out on these clones to determine the genomic location of the seamless vector insertion. Additionally, the strategy should also be tested on a different cell line to identify *attH4X* harboring actively transcribed *LINE-1* elements.

5. CONCLUSION

In summary, a novel λ -Int mediated seamless vector transgenesis platform was developed that can produce as well as site-specifically insert multi-transgene seamless vectors into endogenous sequences present within the *LINE-1* element in the human genome.

In order to generate seamless vector via λ -Integrase mediated intramolecular recombination, both *in vitro* and *in vivo* platforms were developed that provided enough amount of seamless vector for its downstream application. The seamless vector was further validated in HT1080 cells and human embryonic stem cells (hESCs) for λ -Integrase mediated site-specific integration at *attH4X* sequences into human *LINE-1*. Finally, as a proof of concept for therapeutic protein expression, a seamless vector encoding anti-CD19 chimeric antigen receptor (CAR) was inserted and expressed at *LINE-1* in hESCs. The targeted hES clone showed stable long-term expression of anti-CD19 CAR, thus, demonstrating its potential utility in therapeutic application. Furthermore, analyzing all the targeting events in HT1080 cells and hESCs, certain *LINE-1* elements were identified as preferred insertion sites for integrase mediated seamless vector transgenesis.

To explore λ -Int mediated seamless vector transgenesis platform for biologics production, linear ‘landing pads’ carrying 4X *attH4X* sequences along with a selectable marker was randomly inserted in HEK293T cells followed by λ -integrase mediated targeting of the seamless vector at *attH4X* sites of the landing pads. This resulted in the high transgene expressing clones that harbors seamless vector in one of the four *attH4X* sequences of the landing pad. This platform was further utilized to integrate and express full-length factor VIII as a proof of concept for biologics production. The targeted bulk HEK293T culture showed a high production of factor VIII representing its potential application for recombinant protein production in the biopharmaceutical sector. Moreover, this platform also exhibited the ability to site-specifically insert large transgene of size 14kb into the human genome demonstrating its possible application for gene therapy purpose.

Additionally, to enhance the transgene expression inside the human genome, the CTCF binding sites were included in the seamless vector construct. These seamless vectors carrying CTCF sites were further used for λ -Int mediated site-specific insertion in hESCs. Even though the inclusion of CTCF sites in the seamless vector increased the number of transgenes expressing clones and their level of expression, however, it affected the λ -integrase mediated recombination between *attL4X* on the seamless vector with genomic *attH4X* at the *LINE-1* elements. Thus, further studies need to be conducted to understand how CTCF binding site affected the integrase mediated seamless vector transgenesis. In addition, to increase the specificity of λ -Integrase mediated seamless vector transgenesis tool in term of targeting and transcribing transgene using an internal promoter, seamless vector was designed to target *attH4X* site in an actively transcribed *LINE-1 element* in hESCs. The

approach for targeting actively transcribed *LINE-1 element* was found to be specific and provided fewer clones, however, the genomic location of the seamless vector integration could not be determined with different PCR approaches.

Overall, this novel λ -Int mediated seamless vector transgenesis platform is a safe and efficient version of λ -Int mediated genome editing tool that has the inherent advantages of λ -Int mediated targeting in terms of cytotoxicity, genome integrity and capability of choosing the best performing clones from various targeting attempts (1). Thus, all these together, make this platform attractive for utility at an intersection of molecular medicine, gene therapy and systems/synthetic biology for various biomedical applications.

6. FUTURE WORK

The λ -Int mediated seamless vector transgenesis platform has great potential for therapeutic and biotechnological applications. However, further studies and experiments are required to optimize and improve this technology. It would focus on minimization of off-target effects, engineering of mutant integrase variants for better sequence specificity and selection of neutral *LINE-1* elements for expression of the transgene in both undifferentiated and differentiated cellular states. In addition, the following studies can be conducted to further improve the λ -Int mediated seamless vector transgenesis platform:

- a) Generation of an engineered *E. coli* strain to produce Int-C3 or Int-h/218 for *in vivo* production of the seamless vector. The strain could be additionally engineered to produce endonuclease under an inducible system for digestion unwanted bacterial backbone and unrecombined plasmid. Development of such a system might allow us to obtain purified supercoiled seamless vector for downstream applications.
- b) The λ -Int mediated biologics production platform in HEK293T cells can be utilized to study the effect of CTCF binding sites on the transgene expression. This platform can allow us to precisely study the effect of CTCF binding site from a specific genomic locus by integrating the seamless vector carrying CTCF site at the artificially introduced landing pad. Thus, it might reduce the chromatin position effect and help us to understand the effect of CTCF sites over negative control.
- c) The λ -Int mediated seamless vector transgenesis platform can be further improved in terms of safety, by removing the antibiotic-resistant cassette from the integrated seamless vector at the *LINE-1 element*. This can be achieved by the *Cre-loxP* system where the antibiotic-resistant cassette flanked by *loxP* sites will be deleted by Cre recombinase, thus, making the platform suitable for clinical purposes.

REFERENCES

1. Vijaya Chandra, S.H., Makhija, H., Peter, S., Myint Wai, C.M., Li, J., Zhu, J., Ren, Z., D'Alcontres, M.S., Siau, J.W., Chee, S. *et al.* (2016) Conservative site-specific and single-copy transgenesis in human LINE-1 elements. *Nucleic Acids Res*, **44**, e55.
2. Gaj, T., Sirk, S.J., Shui, S.L. and Liu, J. (2016) Genome-Editing Technologies: Principles and Applications. *Cold Spring Harb Perspect Biol*, **8**.
3. Carroll, D. (2014) Genome engineering with targetable nucleases. *Annu Rev Biochem*, **83**, 409-439.
4. Sadelain, M., Papapetrou, E.P. and Bushman, F.D. (2011) Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer*, **12**, 51-58.
5. Esvelt, K.M. and Wang, H.H. (2013) Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol*, **9**, 641.
6. Mahfouz, M.M., Piatek, A. and Stewart, C.N., Jr. (2014) Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. *Plant Biotechnol J*, **12**, 1006-1014.
7. Sorrell, D.A. and Kolb, A.F. (2005) Targeted modification of mammalian genomes. *Biotechnol Adv*, **23**, 431-469.
8. Niidome, T. and Huang, L. (2002) Gene therapy progress and prospects: nonviral vectors. *Gene therapy*, **9**, 1647-1652.
9. Wurtele, H., Little, K.C. and Chartrand, P. (2003) Illegitimate DNA integration in mammalian cells. *Gene therapy*, **10**, 1791-1799.
10. Yew, N.S., Zhao, H., Przybylska, M., Wu, I.H., Tousignant, J.D., Scheule, R.K. and Cheng, S.H. (2002) CpG-depleted plasmid DNA vectors with enhanced safety and long-term gene expression in vivo. *Molecular tabletherapy : the journal of the American Society of Gene Therapy*, **5**, 731-738.
11. Hodges, B.L., Taylor, K.M., Joseph, M.F., Bourgeois, S.A. and Scheule, R.K. (2004) Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides. *Molecular therapy : the journal of the American Society of Gene Therapy*, **10**, 269-278.
12. Meir, Y.J., Lin, A., Huang, M.F., Lin, J.R., Weirauch, M.T., Chou, H.C., Lin, S.J. and Wu, S.C. (2013) A versatile, highly efficient, and potentially safer piggyBac transposon system for mammalian genome manipulations. *FASEB J*, **27**, 4429-4443.
13. Rio, D.C. (1991) Regulation of Drosophila P element transposition. *Trends Genet*, **7**, 282-287.
14. Ivics, Z., Hackett, P.B., Plasterk, R.H. and Izsvak, Z. (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, **91**, 501-510.
15. Hackett, P.B., Ekker, S.C., Largaespada, D.A. and McIvor, R.S. (2005) Sleeping beauty transposon-mediated gene therapy for prolonged expression. *Adv Genet*, **54**, 189-232.
16. Wu, S.C., Meir, Y.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S. and Kaminski, J.M. (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci U S A*, **103**, 15008-15013.
17. Yusa, K., Rad, R., Takeda, J. and Bradley, A. (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nature methods*, **6**, 363-369.
18. Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J. *et al.* (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet*, **41**, 753-761.
19. Furushima, K., Jang, C.W., Chen, D.W., Xiao, N., Overbeek, P.A. and Behringer, R.R. (2012) Insertional mutagenesis by a hybrid piggyBac and sleeping beauty transposon in the rat. *Genetics*, **192**, 1235-1248.
20. de Silva, S., Mastrangelo, M.A., Lotta, L.T., Jr., Burris, C.A., Federoff, H.J. and Bowers, W.J. (2010) Extending the transposable payload limit of Sleeping Beauty (SB) using the Herpes Simplex Virus (HSV)/SB amplicon-vector platform. *Gene Ther*, **17**, 424-431.
21. Di Matteo, M., Samara-Kuko, E., Ward, N.J., Waddington, S.N., McVey, J.H., Chuah, M.K. and VandenDriessche, T. (2014) Hyperactive piggyBac transposons for sustained and robust liver-targeted gene therapy. *Mol Ther*, **22**, 1614-1624.
22. Katter, K., Geurts, A.M., Hoffmann, O., Mates, L., Landa, V., Hiripi, L., Moreno, C., Lazar, J., Bashir, S., Zidek, V. *et al.* (2013) Transposon-mediated transgenesis, transgenic rescue, and tissue-specific gene expression in rodents and rabbits. *FASEB J*, **27**, 930-941.

23. Li, M.A., Turner, D.J., Ning, Z., Yusa, K., Liang, Q., Eckert, S., Rad, L., Fitzgerald, T.W., Craig, N.L. and Bradley, A. (2011) Mobilization of giant piggyBac transposons in the mouse genome. *Nucleic Acids Res*, **39**, e148.
24. Wu H., C.A.K., Mahato R.I. (2019) Gene Therapy. In: Crommelin D., Sindelar R., Meibohm B. (eds). *Pharmaceutical Biotechnology*. Springer, Cham, 323-355.
25. Thomas, C.E., Ehrhardt, A. and Kay, M.A. (2003) Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet*, **4**, 346-358.
26. Howarth, J.L., Lee, Y.B. and Uney, J.B. (2010) Using viral vectors as gene transfer tools (Cell Biology and Toxicology Special Issue: ETCS-UK 1 day meeting on genetic manipulation of cells). *Cell Biology and Toxicology*, **26**, 1-20.
27. Henckaerts, E. and Linden, R.M. (2010) Adeno-associated virus: a key to the human genome? *Future Virol*, **5**, 555-574.
28. Thaci, B., Ulasov, I.V., Wainwright, D.A. and Lesniak, M.S. (2011) The challenge for gene therapy: innate immune response to adenoviruses. *Oncotarget*, **2**, 113-121.
29. Vannucci, L., Lai, M., Chiuppesi, F., Ceccherini-Nelli, L. and Pistello, M. (2013) Viral vectors: a look back and ahead on gene transfer technology. *New Microbiol*, **36**, 1-22.
30. Lundstrom, K. (2018) Viral Vectors in Gene Therapy. *Diseases*, **6**.
31. Samulski, R.J., Zhu, X., Xiao, X., Brook, J.D., Housman, D.E., Epstein, N. and Hunter, L.A. (1991) Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J*, **10**, 3941-3950.
32. Samulski, R.J. and Muzyczka, N. (2014) AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu Rev Virol*, **1**, 427-451.
33. Park, K., Kim, W.J., Cho, Y.H., Lee, Y.I., Lee, H., Jeong, S., Cho, E.S., Chang, S.I., Moon, S.K., Kang, B.S. *et al.* (2008) Cancer gene therapy using adeno-associated virus vectors. *Front Biosci*, **13**, 2653-2659.
34. Schiedner, G., Morral, N., Parks, R.J., Wu, Y., Koopmans, S.C., Langston, C., Graham, F.L., Beaudet, A.L. and Kochanek, S. (1998) Genomic DNA transfer with a high-capacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. *Nat Genet*, **18**, 180-183.
35. Holmes, K.D., Cassam, A.K., Chan, B., Peters, A.A., Weaver, L.C. and Dekaban, G.A. (2000) A multi-mutant herpes simplex virus vector has minimal cytotoxic effects on the distribution of filamentous actin, alpha-actinin 2 and a glutamate receptor in differentiated PC12 cells. *J Neurovirol*, **6**, 33-45.
36. Kay, M.A., Glorioso, J.C. and Naldini, L. (2001) Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med*, **7**, 33-40.
37. Vigna, E. and Naldini, L. (2000) Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. *J Gene Med*, **2**, 308-316.
38. McCarthy, H.O., Wang, Y., Mangipudi, S.S. and Hatefi, A. (2010) Advances with the use of bio-inspired vectors towards creation of artificial viruses. *Expert Opin Drug Deliv*, **7**, 497-512.
39. Strausbaugh, L.J. (2000) Gene therapy and infection control: more light on the way. *Infect Control Hosp Epidemiol*, **21**, 630-632.
40. Sasaki, M. and Plate, K.H. (1998) Gene therapy of malignant glioma: recent advances in experimental and clinical studies. *Ann Oncol*, **9**, 1155-1166.
41. Adamson, A.D., Jackson, D. and Davis, J.R. (2011) Novel approaches to in vitro transgenesis. *J Endocrinol*, **208**, 193-206.
42. Dillon, N. and Festenstein, R. (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet*, **18**, 252-258.
43. Gaj, T., Gersbach, C.A. and Barbas, C.F., 3rd. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, **31**, 397-405.
44. Rouet, P., Smih, F. and Jasin, M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, **14**, 8096-8106.
45. Choulika, A., Perrin, A., Dujon, B. and Nicolas, J.F. (1995) Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol*, **15**, 1968-1973.
46. Bibikova, M., Golic, M., Golic, K.G. and Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, **161**, 1169-1175.
47. Bernheim, A., Calvo-Villamanan, A., Basier, C., Cui, L., Rocha, E.P.C., Touchon, M. and Bikard, D. (2017) Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat Commun*, **8**, 2094.

48. Mao, Z., Bozzella, M., Seluanov, A. and Gorbunova, V. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*, **7**, 2902-2906.
49. Arnoult, N., Correia, A., Ma, J., Merlo, A., Garcia-Gomez, S., Maric, M., Tognetti, M., Benner, C.W., Boulton, S.J., Saghatelian, A. *et al.* (2017) Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature*, **549**, 548-552.
50. Auer, T.O., Duroure, K., De Cian, A., Concordet, J.P. and Del Bene, F. (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res*, **24**, 142-153.
51. Maresca, M., Lin, V.G., Guo, N. and Yang, Y. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res*, **23**, 539-546.
52. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z. *et al.* (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144-149.
53. Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
54. Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol*, **21**, 289-297.
55. Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
56. Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646-651.
57. Lin, S., Staahl, B.T., Alla, R.K. and Doudna, J.A. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*, **3**, e04766.
58. Branzei, D. and Foiani, M. (2008) Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol*, **9**, 297-308.
59. Jacquier, A. and Dujon, B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383-394.
60. Hafez, M. and Hausner, G. (2012) Homing endonucleases: DNA scissors on a mission. *Genome*, **55**, 553-569.
61. Paques, F. and Duchateau, P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther*, **7**, 49-66.
62. Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P. and Paques, F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther*, **11**, 11-27.
63. Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J. and Stoddard, B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell*, **10**, 895-905.
64. Boissel, S., Jarjour, J., Astrakhan, A., Adey, A., Gouble, A., Duchateau, P., Shendure, J., Stoddard, B.L., Certo, M.T., Baker, D. *et al.* (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res*, **42**, 2591-2601.
65. Delacote, F., Perez, C., Guyot, V., Duhamel, M., Rochon, C., Ollivier, N., Macmaster, R., Silva, G.H., Paques, F., Daboussi, F. *et al.* (2013) High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes. *PLoS One*, **8**, e53217.
66. Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 1156-1160.
67. Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S. and Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic acids research*, **28**, 3361-3369.
68. Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809-817.
69. Santiago, Y., Chan, E., Liu, P.Q., Orlando, S., Zhang, L., Urnov, F.D., Holmes, M.C., Guschin, D., Waite, A., Miller, J.C. *et al.* (2008) Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A*, **105**, 5809-5814.

70. Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKolver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol*, **27**, 851-857.
71. Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J. *et al.* (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol*, **26**, 702-708.
72. Capecchi, M.R. (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet*, **6**, 507-512.
73. Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L. *et al.* (2011) In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, **475**, 217-221.
74. Genovese, P., Schirotti, G., Escobar, G., Tomaso, T.D., Firrito, C., Calabria, A., Moi, D., Mazziere, R., Bonini, C., Holmes, M.C. *et al.* (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*, **510**, 235-240.
75. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature methods*, **8**, 765-770.
76. Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppl, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology*, **29**, 816-823.
77. Cox, D.B., Platt, R.J. and Zhang, F. (2015) Therapeutic genome editing: prospects and challenges. *Nat Med*, **21**, 121-131.
78. Kandavelou, K. and Chandrasegaran, S. (2009) Custom-designed molecular scissors for site-specific manipulation of the plant and mammalian genomes. *Methods Mol Biol*, **544**, 617-636.
79. Hui Zhang, J.Z., Zhaobo Lang, José Ramón Botella & Jian-Kang Zhu. (2017) Genome Editing—Principles and Applications for Functional Genomics Research and Crop Improvement. *Critical Reviews in Plant Sciences*, **36**, 291-309.
80. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757-761.
81. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, **29**, 143-148.
82. Mussolino, C., Alzubi, J., Fine, E.J., Morbitzer, R., Cradick, T.J., Lahaye, T., Bao, G. and Cathomen, T. (2014) TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic acids research*, **42**, 6762-6773.
83. Wienert, B., Funnell, A.P., Norton, L.J., Pearson, R.C., Wilkinson-White, L.E., Lester, K., Vadolas, J., Porteus, M.H., Matthews, J.M., Quinlan, K.G. *et al.* (2015) Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat Commun*, **6**, 7085.
84. Ramalingam, S., Annaluru, N., Kandavelou, K. and Chandrasegaran, S. (2014) TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. *Curr Gene Ther*, **14**, 461-472.
85. Osborn, M.J., Starker, C.G., McElroy, A.N., Webber, B.R., Riddle, M.J., Xia, L., DeFeo, A.P., Gabriel, R., Schmidt, M., von Kalle, C. *et al.* (2013) TALEN-based gene correction for epidermolysis bullosa. *Mol Ther*, **21**, 1151-1159.
86. Dupuy, A., Valton, J., Leduc, S., Armier, J., Galetto, R., Gouble, A., Lebuhotel, C., Stary, A., Paques, F., Duchateau, P. *et al.* (2013) Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN. *PLoS One*, **8**, e78678.
87. Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.
88. Davies, B., Davies, G., Preece, C., Puliyadi, R., Szumska, D. and Bhattacharya, S. (2013) Site specific mutation of the Zic2 locus by microinjection of TALEN mRNA in mouse CD1, C3H and C57BL/6J oocytes. *PLoS One*, **8**, e60216.
89. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.

90. Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*, **1**, e60.
91. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.
92. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols*, **8**, 2281-2308.
93. Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology*, **32**, 347-355.
94. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R. *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523**, 481-485.
95. Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat Commun*, **9**, 1911.
96. Guha, T.K., Wai, A. and Hausner, G. (2017) Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Comput Struct Biotechnol J*, **15**, 146-160.
97. Zhang, H.X., Zhang, Y. and Yin, H. (2019) Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Mol Ther*, **27**, 735-746.
98. Ho, B.X., Loh, S.J.H., Chan, W.K. and Soh, B.S. (2018) In Vivo Genome Editing as a Therapeutic Approach. *Int J Mol Sci*, **19**.
99. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
100. Chiang, T.W., le Sage, C., Larrieu, D., Demir, M. and Jackson, S.P. (2016) CRISPR-Cas9(D10A) nickase-based genotypic and phenotypic screening to enhance genome editing. *Sci Rep*, **6**, 24356.
101. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380-1389.
102. Jensen, M.K. (2018) Design principles for nuclease-deficient CRISPR-based transcriptional regulators. *FEMS Yeast Res*, **18**.
103. Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S. *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, **155**, 1479-1491.
104. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149-157.
105. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I. and Liu, D.R. (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature*, **551**, 464-471.
106. Eid, A., Alshareef, S. and Mahfouz, M.M. (2018) CRISPR base editors: genome editing without double-stranded breaks. *Biochem J*, **475**, 1955-1964.
107. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420-424.
108. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759-771.
109. Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N. and Zhang, F. (2017) Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol*, **35**, 789-792.
110. Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J. and Joung, J.K. (2016) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol*, **34**, 869-874.
111. Safari, F., Zare, K., Negahdaripour, M., Barekati-Mowahed, M. and Ghasemi, Y. (2019) CRISPR Cpf1 proteins: structure, function and implications for genome editing. *Cell Biosci*, **9**, 36.
112. Bayat, H., Modarressi, M.H. and Rahimpour, A. (2018) The Conspicuity of CRISPR-Cpf1 System as a Significant Breakthrough in Genome Editing. *Curr Microbiol*, **75**, 107-115.

113. Portolano, N., Watson, P.J., Fairall, L., Millard, C.J., Milano, C.P., Song, Y., Cowley, S.M. and Schwabe, J.W. (2014) Recombinant protein expression for structural biology in HEK 293F suspension cells: a novel and accessible approach. *J Vis Exp*, e51897.
114. Moon, S.B., Kim, D.Y., Ko, J.H. and Kim, Y.S. (2019) Recent advances in the CRISPR genome editing tool set. *Exp Mol Med*, **51**, 130.
115. Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kuhn, R. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology*, **33**, 543-548.
116. Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B. and Vakoc, C.R. (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nature biotechnology*, **33**, 661-667.
117. Shechner, D.M., Hacisuleyman, E., Younger, S.T. and Rinn, J.L. (2015) Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nature methods*, **12**, 664-670.
118. Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E, P.R.I., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J. *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nature methods*, **12**, 326-328.
119. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S. and Vale, R.D. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, **159**, 635-646.
120. Liu, Y., Zhan, Y., Chen, Z., He, A., Li, J., Wu, H., Liu, L., Zhuang, C., Lin, J., Guo, X. *et al.* (2016) Directing cellular information flow via CRISPR signal conductors. *Nature methods*, **13**, 938-944.
121. Freedman, B.S., Brooks, C.R., Lam, A.Q., Fu, H., Morizane, R., Agrawal, V., Saad, A.F., Li, M.K., Hughes, M.R., Werff, R.V. *et al.* (2015) Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat Commun*, **6**, 8715.
122. Bellin, M., Casini, S., Davis, R.P., D'Aniello, C., Haas, J., Ward-van Oostwaard, D., Tertoolen, L.G., Jung, C.B., Elliott, D.A., Welling, A. *et al.* (2013) Isogenic human pluripotent stem cell pairs reveal the role of a KCNH2 mutation in long-QT syndrome. *EMBO J*, **32**, 3161-3175.
123. Shalem, O., Sanjana, N.E. and Zhang, F. (2015) High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet*, **16**, 299-311.
124. Pellagatti, A., Dolatshad, H., Valletta, S. and Boulton, J. (2015) Application of CRISPR/Cas9 genome editing to the study and treatment of disease. *Arch Toxicol*, **89**, 1023-1034.
125. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, **31**, 822-826.
126. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
127. Byrne, S.M., Ortiz, L., Mali, P., Aach, J. and Church, G.M. (2015) Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic acids research*, **43**, e21.
128. Swiech, K., Picanco-Castro, V. and Covas, D.T. (2017) Production of recombinant coagulation factors: Are humans the best host cells? *Bioengineered*, **8**, 462-470.
129. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V. and Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, **365**, 48-53.
130. Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S. and Sternberg, S.H. (2019) Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature*, **571**, 219-225.
131. Kwon, J.B. and Gersbach, C.A. (2019) Jumping at the chance for precise DNA integration. *Nat Biotechnol*, **37**, 1004-1006.
132. Smith, K. (2001) Theoretical mechanisms in targeted and random integration of transgene DNA. *Reprod Nutr Dev*, **41**, 465-485.
133. Vasquez, K.M., Marburger, K., Intody, Z. and Wilson, J.H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci U S A*, **98**, 8403-8410.
134. Akopian, A. and Marshall Stark, W. (2005) Site-specific DNA recombinases as instruments for genomic surgery. *Adv Genet*, **55**, 1-23.
135. NASH, H.A. (1996) Site-specific recombination: Integration, excision, resolution, and inversion of defined DNA segments. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2363-2376.
136. Olorunniji, F.J., Rosser, S.J. and Stark, W.M. (2016) Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem J*, **473**, 673-684.
137. Grindley, N.D., Whiteson, K.L. and Rice, P.A. (2006) Mechanisms of site-specific recombination. *Annual review of biochemistry*, **75**, 567-605.

138. Holliday, R. (2007) A mechanism for gene conversion in fungi. *Genet Res*, **89**, 285-307.
139. Kim, A.I., Ghosh, P., Aaron, M.A., Bibb, L.A., Jain, S. and Hatfull, G.F. (2003) Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol Microbiol*, **50**, 463-473.
140. Ghosh, P., Wasil, L.R. and Hatfull, G.F. (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol*, **4**, e186.
141. Singh, S., Ghosh, P. and Hatfull, G.F. (2013) Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. *PLoS Genet*, **9**, e1003490.
142. Russell, J.P., Chang, D.W., Tretiakova, A. and Padidam, M. (2006) Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *Biotechniques*, **40**, 460, 462, 464.
143. Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M. and Brown, W. (2013) Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol*, **13**, 87.
144. Inniss, M.C., Bandara, K., Jusiak, B., Lu, T.K., Weiss, R., Wroblewska, L. and Zhang, L. (2017) A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnol Bioeng*, **114**, 1837-1846.
145. Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W.A., Lowe, A. *et al.* (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res*, **46**, 4072-4086.
146. Smith, M.C. and Thorpe, H.M. (2002) Diversity in the serine recombinases. *Mol Microbiol*, **44**, 299-307.
147. Groth, A.C. and Calos, M.P. (2004) Phage integrases: biology and applications. *J Mol Biol*, **335**, 667-678.
148. Hirano, N., Muroi, T., Takahashi, H. and Haruki, M. (2011) Site-specific recombinases as tools for heterologous gene integration. *Appl Microbiol Biotechnol*, **92**, 227-239.
149. Brown, W.R., Lee, N.C., Xu, Z. and Smith, M.C. (2011) Serine recombinases as tools for genome engineering. *Methods*, **53**, 372-379.
150. Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S. and Calos, M.P. (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol*, **21**, 3926-3934.
151. Chalberg, T.W., Genise, H.L., Vollrath, D. and Calos, M.P. (2005) phiC31 integrase confers genomic integration and long-term transgene expression in rat retina. *Investigative ophthalmology & visual science*, **46**, 2140-2146.
152. Liu, J., Skjorring, T., Gjetting, T. and Jensen, T.G. (2009) PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. *BMC biotechnology*, **9**, 31.
153. Austin, S., Ziese, M. and Sternberg, N. (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell*, **25**, 729-736.
154. Broach, J.R., Guarascio, V.R. and Jayaram, M. (1982) Recombination within the yeast plasmid 2mu circle is site-specific. *Cell*, **29**, 227-234.
155. Nagy, A. (2000) Cre recombinase: the universal reagent for genome tailoring. *Genesis*, **26**, 99-109.
156. Sauer, B. and Henderson, N. (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 5166-5170.
157. Utomo, A.R., Nikitin, A.Y. and Lee, W.H. (1999) Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice. *Nature biotechnology*, **17**, 1091-1096.
158. Livet, J., Weissman, T.A., Kang, H., Draft, R.W., Lu, J., Bennis, R.A., Sanes, J.R. and Lichtman, J.W. (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, **450**, 56-62.
159. Callesen, M.M., Berthelsen, M.F., Lund, S., Fuchtbauer, A.C., Fuchtbauer, E.M. and Jakobsen, J.E. (2016) Recombinase-Mediated Cassette Exchange (RMCE)-in Reporter Cell Lines as an Alternative to the FLP-in System. *PLoS One*, **11**, e0161471.
160. Turan, S., Galla, M., Ernst, E., Qiao, J., Voelkel, C., Schiedlmeier, B., Zehe, C. and Bode, J. (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J Mol Biol*, **407**, 193-221.
161. Loonstra, A., Vooijs, M., Beverloo, H.B., Allak, B.A., van Drunen, E., Kanaar, R., Berns, A. and Jonkers, J. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci U S A*, **98**, 9209-9214.

162. Schebelle, L., Wolf, C., Stribl, C., Javaheri, T., Schnutgen, F., Ettinger, A., Ivics, Z., Hansen, J., Ruiz, P., von Melchner, H. *et al.* (2010) Efficient conditional and promoter-specific in vivo expression of cDNAs of choice by taking advantage of recombinase-mediated cassette exchange using FLEX gene traps. *Nucleic Acids Res*, **38**, e106.
163. Dallas, S.L., Xie, Y., Shiflett, L.A. and Ueki, Y. (2018) Mouse Cre Models for the Study of Bone Diseases. *Curr Osteoporos Rep*, **16**, 466-477.
164. Landy, A. (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem*, **58**, 913-949.
165. Fogg, P.C., Colloms, S., Rosser, S., Stark, M. and Smith, M.C. (2014) New applications for phage integrases. *J Mol Biol*, **426**, 2703-2716.
166. Nunes-Duby, S.E., Kwon, H.J., Tirumalai, R.S., Ellenberger, T. and Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic acids research*, **26**, 391-406.
167. Landy, A. (2015) The lambda Integrase Site-specific Recombination Pathway. *Microbiol Spectr*, **3**, MDNA3-0051-2014.
168. Papagiannis, C.V., Sam, M.D., Abbani, M.A., Yoo, D., Cascio, D., Clubb, R.T. and Johnson, R.C. (2007) Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. *Journal of molecular biology*, **367**, 328-343.
169. Richet, E., Abcarian, P. and Nash, H.A. (1988) Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell*, **52**, 9-17.
170. Mizuuchi, K., Gellert, M., Weisberg, R.A. and Nash, H.A. (1980) Catenation and supercoiling in the products of bacteriophage lambda integrative recombination in vitro. *Journal of molecular biology*, **141**, 485-494.
171. Burgin, A.B., Jr. and Nash, H.A. (1992) Symmetry in the mechanism of bacteriophage lambda integrative recombination. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 9642-9646.
172. Nunes-Duby, S.E., Matsumoto, L. and Landy, A. (1987) Site-specific recombination intermediates trapped with suicide substrates. *Cell*, **50**, 779-788.
173. Christ, N. and Droge, P. (1999) Alterations in the directionality of lambda site-specific recombination catalyzed by mutant integrases in vivo. *J Mol Biol*, **288**, 825-836.
174. Landy, A. (1993) Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr Opin Genet Dev*, **3**, 699-707.
175. Yang, W. and Mizuuchi, K. (1997) Site-specific recombination in plane view. *Structure*, **5**, 1401-1406.
176. Praznovszky, T. (2011) Chromosome engineering with lambda-integrase mediated recombination system: the ACE system. *Methods Mol Biol*, **738**, 141-149.
177. Miller, H.I., Mozola, M.A. and Friedman, D.I. (1980) int-h: An int mutation of phage lambda that enhances site-specific recombination. *Cell*, **20**, 721-729.
178. Lorbach, E., Christ, N., Schwikardi, M. and Droge, P. (2000) Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. *J Mol Biol*, **296**, 1175-1181.
179. Tay, Y., Ho, C., Droge, P. and Ghadessy, F.J. (2010) Selection of bacteriophage lambda integrases with altered recombination specificity by in vitro compartmentalization. *Nucleic Acids Res*, **38**, e25.
180. Siau, J.W., Chee, S., Makhija, H., Wai, C.M., Chandra, S.H., Peter, S., Droge, P. and Ghadessy, F.J. (2015) Directed evolution of lambda integrase activity and specificity by genetic derepression. *Protein Eng Des Sel*, **28**, 211-220.
181. Makhija, H., Roy, S., Hoon, S., Ghadessy, F.J., Wong, D., Jaiswal, R., Campana, D. and Droge, P. (2018) A novel lambda integrase-mediated seamless vector transgenesis platform for therapeutic protein expression. *Nucleic Acids Res*.
182. Jechlinger, W. (2006) Optimization and delivery of plasmid DNA for vaccination. *Expert review of vaccines*, **5**, 803-825.
183. Chen, Z.Y., He, C.Y., Meuse, L. and Kay, M.A. (2004) Silencing of episomal transgene expression by plasmid bacterial DNA elements in vivo. *Gene therapy*, **11**, 856-864.
184. Suzuki, M., Kasai, K. and Saeki, Y. (2006) Plasmid DNA sequences present in conventional herpes simplex virus amplicon vectors cause rapid transgene silencing by forming inactive chromatin. *J Virol*, **80**, 3293-3300.
185. Riu, E., Chen, Z.Y., Xu, H., He, C.Y. and Kay, M.A. (2007) Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. *Molecular therapy : the journal of the American Society of Gene Therapy*, **15**, 1348-1355.

186. Klinman, D.M. (2004) Immunotherapeutic uses of CpG oligodeoxynucleotides. *Nature reviews. Immunology*, **4**, 249-258.
187. Krieg, A.M., Yi, A.K., Matson, S., Waldschmidt, T.J., Bishop, G.A., Teasdale, R., Koretzky, G.A. and Klinman, D.M. (1995) CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature*, **374**, 546-549.
188. Kreiss, P., Cameron, B., Rangara, R., Mailhe, P., Aguerre-Charriol, O., Airiau, M., Scherman, D., Crouzet, J. and Pitard, B. (1999) Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic acids research*, **27**, 3792-3798.
189. Yin, W., Xiang, P. and Li, Q. (2005) Investigations of the effect of DNA size in transient transfection assay using dual luciferase system. *Anal Biochem*, **346**, 289-294.
190. Darquet, A.M., Rangara, R., Kreiss, P., Schwartz, B., Naimi, S., Delaere, P., Crouzet, J. and Scherman, D. (1999) Minicircle: an improved DNA molecule for in vitro and in vivo gene transfer. *Gene therapy*, **6**, 209-218.
191. Mayrhofer, P., Schleef, M. and Jechlinger, W. (2009) Use of minicircle plasmids for gene therapy. *Methods Mol Biol*, **542**, 87-104.
192. Darquet, A.M., Cameron, B., Wils, P., Scherman, D. and Crouzet, J. (1997) A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. *Gene Ther*, **4**, 1341-1349.
193. Kreiss, P., Cameron, B., Darquet, A.M., Scherman, D. and Crouzet, J. (1998) Production of a new DNA vehicle for gene transfer using site-specific recombination. *Appl Microbiol Biotechnol*, **49**, 560-567.
194. Chen, Z.Y., He, C.Y., Ehrhardt, A. and Kay, M.A. (2003) Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo. *Mol Ther*, **8**, 495-500.
195. Bigger, B.W., Tolmachov, O., Collombet, J.M., Fragkos, M., Palaszewski, I. and Coutelle, C. (2001) An araC-controlled bacterial cre expression system to produce DNA minicircle vectors for nuclear and mitochondrial gene therapy. *J Biol Chem*, **276**, 23018-23027.
196. Jechlinger, W., Azimpour Tabrizi, C., Lubitz, W. and Mayrhofer, P. (2004) Minicircle DNA immobilized in bacterial ghosts: in vivo production of safe non-viral DNA delivery vehicles. *J Mol Microbiol Biotechnol*, **8**, 222-231.
197. Gilbertson, L. (2003) Cre-lox recombination: Cre-ative tools for plant biotechnology. *Trends Biotechnol*, **21**, 550-555.
198. Sadowski, P. (1986) Site-specific recombinases: changing partners and doing the twist. *J Bacteriol*, **165**, 341-347.
199. Thomson, J.G. and Ow, D.W. (2006) Site-specific recombination systems for the genetic manipulation of eukaryotic genomes. *Genesis*, **44**, 465-476.
200. Eberl, L., Kristensen, C.S., Givskov, M., Grohmann, E., Gerlitz, M. and Schwab, H. (1994) Analysis of the multimer resolution system encoded by the parCBA operon of broad-host-range plasmid RP4. *Mol Microbiol*, **12**, 131-141.
201. Mayrhofer, P., Blaesen, M., Schleef, M. and Jechlinger, W. (2008) Minicircle-DNA production by site specific recombination and protein-DNA interaction chromatography. *J Gene Med*, **10**, 1253-1269.
202. Bandaranayake, A.D. and Almo, S.C. (2014) Recent advances in mammalian protein production. *FEBS Lett*, **588**, 253-260.
203. Butler, M. and Meneses-Acosta, A. (2012) Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Appl Microbiol Biotechnol*, **96**, 885-894.
204. Tabandeh, F., Shojaosadati, S.A., Zomorodipour, A., Khodabandeh, M., Sanati, M.H. and Yakhchali, B. (2004) Heat-induced production of human growth hormone by high cell density cultivation of recombinant Escherichia coli. *Biotechnol Lett*, **26**, 245-250.
205. Majidzadeh, A.K., Khalaj, V., Fatemeh, D., Mahdi, H., Farzaneh, B., Ahmad, A. and Mahboudi, F. (2010) Cloning and expression of functional full-length human tissue plasminogen activator in Pichia pastoris. *Appl Biochem Biotechnol*, **162**, 2037-2048.
206. Ahmad, M., Hirz, M., Pichler, H. and Schwab, H. (2014) Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol*, **98**, 5301-5317.
207. Zhu, J. (2012) Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv*, **30**, 1158-1170.
208. Dalton, A.C. and Barton, W.A. (2014) Over-expression of secreted proteins from mammalian cell lines. *Protein Sci*, **23**, 517-525.

209. Hacker, D.L. and Balasubramanian, S. (2016) Recombinant protein production from stable mammalian cell lines and pools. *Curr Opin Struct Biol*, **38**, 129-136.
210. Chu, L. and Robinson, D.K. (2001) Industrial choices for protein production by large-scale cell culture. *Curr Opin Biotechnol*, **12**, 180-187.
211. Birch, J.R. and Onakunle, Y. (2005) Biopharmaceutical proteins: opportunities and challenges. *Methods Mol Biol*, **308**, 1-16.
212. Jazayeri, S.H., Amiri-Yekta, A., Bahrami, S., Gourabi, H., Sanati, M.H. and Khorramizadeh, M.R. (2018) Vector and Cell Line Engineering Technologies Toward Recombinant Protein Expression in Mammalian Cell Lines. *Appl Biochem Biotechnol*, **185**, 986-1003.
213. Li, Z., Michael, I.P., Zhou, D., Nagy, A. and Rini, J.M. (2013) Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. *Proc Natl Acad Sci U S A*, **110**, 5004-5009.
214. Caro, L.N., Li, Z., Balo, A.R., Van Eps, N., Rini, J.M. and Ernst, O.P. (2015) Rapid and facile recombinant expression of bovine rhodopsin in HEK293S GnTI(-) cells using a PiggyBac inducible system. *Methods Enzymol*, **556**, 307-330.
215. Balasubramanian, S., Rajendra, Y., Baldi, L., Hacker, D.L. and Wurm, F.M. (2016) Comparison of three transposons for the generation of highly productive recombinant CHO cell pools and cell lines. *Biotechnol Bioeng*, **113**, 1234-1243.
216. Gaillet, B., Gilbert, R., Broussau, S., Pilotte, A., Malenfant, F., Mullick, A., Garnier, A. and Massie, B. (2010) High-level recombinant protein production in CHO cells using lentiviral vectors and the cumate gene-switch. *Biotechnol Bioeng*, **106**, 203-215.
217. Oberbek, A., Matasci, M., Hacker, D.L. and Wurm, F.M. (2011) Generation of stable, high-producing CHO cell lines by lentiviral vector-mediated gene transfer in serum-free suspension culture. *Biotechnol Bioeng*, **108**, 600-610.
218. Moiani, A., Paleari, Y., Sartori, D., Mezzadra, R., Miccio, A., Cattoglio, C., Cocchiarella, F., Lidonnici, M.R., Ferrari, G. and Mavilio, F. (2012) Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest*, **122**, 1653-1666.
219. Mao, Y., Yan, R., Li, A., Zhang, Y., Li, J., Du, H., Chen, B., Wei, W., Zhang, Y., Summers, C. *et al.* (2015) Lentiviral Vectors Mediate Long-Term and High Efficiency Transgene Expression in HEK 293T cells. *Int J Med Sci*, **12**, 407-415.
220. Baranyi, L., Doering, C.B., Denning, G., Gautney, R.E., Harris, K.T., Spencer, H.T., Roy, A., Zayed, H. and Dropulic, B. (2013) Rapid generation of stable cell lines expressing high levels of erythropoietin, factor VIII, and an antihuman CD20 antibody using lentiviral vectors. *Hum Gene Ther Methods*, **24**, 214-227.
221. Goetze, S., Baer, A., Winkelmann, S., Nehlsen, K., Seibler, J., Maass, K. and Bode, J. (2005) Performance of genomic bordering elements at predefined genomic loci. *Mol Cell Biol*, **25**, 2260-2272.
222. Lai, T., Yang, Y. and Ng, S.K. (2013) Advances in Mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals (Basel)*, **6**, 579-603.
223. Turan, S., Zehe, C., Kuehle, J., Qiao, J. and Bode, J. (2013) Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene*, **515**, 1-27.
224. Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R. and Hauser, H. (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr Opin Biotechnol*, **18**, 411-419.
225. Merrick, C.A., Zhao, J. and Rosser, S.J. (2018) Serine Integrases: Advancing Synthetic Biology. *ACS Synth Biol*, **7**, 299-310.
226. Lee, J.S., Kallehauge, T.B., Pedersen, L.E. and Kildegaard, H.F. (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci Rep*, **5**, 8572.
227. Kim, H. and Kim, J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat Rev Genet*, **15**, 321-334.
228. Lee, J.S., Grav, L.M., Pedersen, L.E., Lee, G.M. and Kildegaard, H.F. (2016) Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol Bioeng*, **113**, 2518-2523.
229. Phan, Q.V., Contzen, J., Seemann, P. and Gossen, M. (2017) Site-specific chromosomal gene insertion: FLP recombinase versus Cas9 nuclease. *Sci Rep*, **7**, 17771.

230. Robertson, G., Garrick, D., Wu, W., Kearns, M., Martin, D. and Whitelaw, E. (1995) Position-dependent variegation of globin transgene expression in mice. *Proc Natl Acad Sci U S A*, **92**, 5371-5375.
231. Wakimoto, B.T. (1998) Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*. *Cell*, **93**, 321-324.
232. Rivella, S. and Sadelain, M. (1998) Genetic treatment of severe hemoglobinopathies: the combat against transgene variegation and transgene silencing. *Semin Hematol*, **35**, 112-125.
233. Mutskov, V.J., Farrell, C.M., Wade, P.A., Wolffe, A.P. and Felsenfeld, G. (2002) The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev*, **16**, 1540-1554.
234. Wong, S.P., Argyros, O. and Harbottle, R.P. (2015) Sustained expression from DNA vectors. *Adv Genet*, **89**, 113-152.
235. Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, **13**, 2465-2477.
236. Li, Q., Peterson, K.R., Fang, X. and Stamatoyannopoulos, G. (2002) Locus control regions. *Blood*, **100**, 3077-3086.
237. Festenstein, R. and Kioussis, D. (2000) Locus control regions and epigenetic chromatin modifiers. *Curr Opin Genet Dev*, **10**, 199-203.
238. Miao, C.H., Ye, X. and Thompson, A.R. (2003) High-level factor VIII gene expression in vivo achieved by nonviral liver-specific gene therapy vectors. *Hum Gene Ther*, **14**, 1297-1305.
239. Recillas-Targa, F., Valadez-Graham, V. and Farrell, C.M. (2004) Prospects and implications of using chromatin insulators in gene therapy and transgenesis. *Bioessays*, **26**, 796-807.
240. West, A.G., Gaszner, M. and Felsenfeld, G. (2002) Insulators: many functions, many mechanisms. *Genes Dev*, **16**, 271-288.
241. Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. (2002) The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A*, **99 Suppl 4**, 16433-16437.
242. Litt, M.D., Simpson, M., Recillas-Targa, F., Prioleau, M.N. and Felsenfeld, G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J*, **20**, 2224-2235.
243. Neville, J.J., Orlando, J., Mann, K., McCloskey, B. and Antoniou, M.N. (2017) Ubiquitous Chromatin-opening Elements (UCOE): Applications in biomanufacturing and gene therapy. *Biotechnol Adv*, **35**, 557-564.
244. Hou, J.J., Hughes, B.S., Smede, M., Leung, K.M., Levine, K., Rigby, S., Gray, P.P. and Munro, T.P. (2014) High-throughput ClonePix FL analysis of mAb-expressing clones using the UCOE expression system. *N Biotechnol*, **31**, 214-220.
245. Nair, A.R., Jinger, X. and Hermiston, T.W. (2011) Effect of different UCOE-promoter combinations in creation of engineered cell lines for the production of Factor VIII. *BMC Res Notes*, **4**, 178.
246. Williams, S., Mustoe, T., Mulcahy, T., Griffiths, M., Simpson, D., Antoniou, M., Irvine, A., Mountain, A. and Crombie, R. (2005) CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol*, **5**, 17.
247. Betts, Z., Croxford, A.S. and Dickson, A.J. (2015) Evaluating the interaction between UCOE and DHFR-linked amplification and stability of recombinant protein expression. *Biotechnol Prog*, **31**, 1014-1025.
248. Senigl, F., Plachy, J. and Hejnar, J. (2008) The core element of a CpG island protects avian sarcoma and leukemia virus-derived vectors from transcriptional silencing. *J Virol*, **82**, 7818-7827.
249. Zhao, C.P., Guo, X., Chen, S.J., Li, C.Z., Yang, Y., Zhang, J.H., Chen, S.N., Jia, Y.L. and Wang, T.Y. (2017) Matrix attachment region combinations increase transgene expression in transfected Chinese hamster ovary cells. *Sci Rep*, **7**, 42805.
250. Nowak, W., Gawłowska, M., Jarmolowski, A. and Augustyniak, J. (2001) Effect of nuclear matrix attachment regions on transgene expression in tobacco plants. *Acta Biochim Pol*, **48**, 637-646.
251. Hart, C.M. and Laemmli, U.K. (1998) Facilitation of chromatin dynamics by SARs. *Curr Opin Genet Dev*, **8**, 519-525.
252. Cremer, T., Kreth, G., Koester, H., Fink, R.H., Heintzmann, R., Cremer, M., Solovei, I., Zink, D. and Cremer, C. (2000) Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit Rev Eukaryot Gene Expr*, **10**, 179-212.

253. Razin, S.V. (2001) The nuclear matrix and chromosomal DNA loops: is there any correlation between partitioning of the genome into loops and functional domains? *Cell Mol Biol Lett*, **6**, 59-69.
254. Jackson, D.A. and Cook, P.R. (1995) The structural basis of nuclear function. *Int Rev Cytol*, **162A**, 125-149.
255. Jenke, B.H., Fetzer, C.P., Stehle, I.M., Jonsson, F., Fackelmayer, F.O., Conradt, H., Bode, J. and Lipps, H.J. (2002) An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo. *EMBO Rep*, **3**, 349-354.
256. Lipps, H.J. and Bode, J. (2001) Exploiting chromosomal and viral strategies: the design of safe and efficient non-viral gene transfer systems. *Curr Opin Mol Ther*, **3**, 133-141.
257. Bode, J., Benham, C., Knopp, A. and Mielke, C. (2000) Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit Rev Eukaryot Gene Expr*, **10**, 73-90.
258. Li, Q., Zhao, C.P., Lin, Y., Song, C., Wang, F. and Wang, T.Y. (2019) Two human MARs effectively increase transgene expression in transfected CHO cells. *J Cell Mol Med*, **23**, 1613-1616.
259. Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A. and Hadjur, S. (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*, **10**, 1297-1309.
260. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665-1680.
261. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376-380.
262. Dumevska, B., Chami, O., McKernan, R., Goel, D. and Schmidt, U. (2016) Derivation of Genea047 human embryonic stem cell line. *Stem Cell Res*, **16**, 322-326.
263. Benedict, W.F., Weissman, B.E., Mark, C. and Stanbridge, E.J. (1984) Tumorigenicity of human HT1080 fibrosarcoma X normal fibroblast hybrids: chromosome dosage dependency. *Cancer Res*, **44**, 3471-3479.
264. Rasheed, S., Nelson-Rees, W.A., Toth, E.M., Arnstein, P. and Gardner, M.B. (1974) Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer*, **33**, 1027-1033.
265. DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H. and Calos, M.P. (1987) Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol*, **7**, 379-387.
266. Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*, **36**, 59-74.
267. Jessop, L., Bankhead, T., Wong, D. and Segall, A.M. (2000) The amino terminus of bacteriophage lambda integrase is involved in protein-protein interactions during recombination. *J Bacteriol*, **182**, 1024-1034.
268. Corona, T., Bao, Q., Christ, N., Schwartz, T., Li, J. and Droge, P. (2003) Activation of site-specific DNA integration in human cells by a single chain integration host factor. *Nucleic Acids Res*, **31**, 5140-5148.
269. Bliska, J.B. and Cozzarelli, N.R. (1987) Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J Mol Biol*, **194**, 205-218.
270. Nash, H.A. (1996) The HU and IHF Proteins: Accessory Factors for Complex Protein-DNA Assemblies. In: *Regulation of Gene Expression in Escherichia coli.*, 149-179.
271. Azaro M.A. and Landy, A. (2002) λ Int and the λ Int family. *Mobile DNA II. ASM Press*, 118-148.
272. Zechiedrich, E.L., Khodursky, A.B. and Cozzarelli, N.R. (1997) Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in Escherichia coli. *Genes Dev*, **11**, 2580-2592.
273. Zechiedrich, E.L., Khodursky, A.B., Bachellier, S., Schneider, R., Chen, D., Lilley, D.M. and Cozzarelli, N.R. (2000) Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli. *J Biol Chem*, **275**, 8103-8113.
274. Fogg, J.M., Kolmakova, N., Rees, I., Magonov, S., Hansma, H., Perona, J.J. and Zechiedrich, E.L. (2006) Exploring writhe in supercoiled minicircle DNA. *J Phys Condens Matter*, **18**, S145-S159.
275. Zhu, Z. and Huangfu, D. (2013) Human pluripotent stem cells: an emerging model in developmental biology. *Development*, **140**, 705-717.
276. Simcikova, M., Prather, K.L., Prazeres, D.M. and Monteiro, G.A. (2015) Towards effective non-viral gene delivery vector. *Biotechnol Genet Eng Rev*, **31**, 82-107.

277. Huang, M., Chen, Z., Hu, S., Jia, F., Li, Z., Hoyt, G., Robbins, R.C., Kay, M.A. and Wu, J.C. (2009) Novel minicircle vector for gene therapy in murine myocardial infarction. *Circulation*, **120**, S230-237.
278. Stenler, S., Andersson, A., Simonson, O.E., Lundin, K.E., Chen, Z.Y., Kay, M.A., Smith, C.I., Sylven, C. and Blomberg, P. (2009) Gene transfer to mouse heart and skeletal muscles using a minicircle expressing human vascular endothelial growth factor. *J Cardiovasc Pharmacol*, **53**, 18-23.
279. Dietz, W.M., Skinner, N.E., Hamilton, S.E., Jund, M.D., Heitfeld, S.M., Litterman, A.J., Hwu, P., Chen, Z.Y., Salazar, A.M., Ohlfest, J.R. *et al.* (2013) Minicircle DNA is superior to plasmid DNA in eliciting antigen-specific CD8⁺ T-cell responses. *Molecular therapy : the journal of the American Society of Gene Therapy*, **21**, 1526-1535.
280. Wu, J., Xiao, X., Zhao, P., Xue, G., Zhu, Y., Zhu, X., Zheng, L., Zeng, Y. and Huang, W. (2006) Minicircle-IFN γ induces antiproliferative and antitumoral effects in human nasopharyngeal carcinoma. *Clin Cancer Res*, **12**, 4702-4713.
281. Hudecek, M., Gogishvili, T., Monjezi, R., Wegner, J., Shankar, R., Kruesemann, C., Miskey, C., Ivics, Z., Schmeer, M. and Schleef, M. (2016) Minicircle-Based Engineering of Chimeric Antigen Receptor (CAR) T Cells. *Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer*, **209**, 37-50.
282. Zhang, C., Liu, J., Zhong, J.F. and Zhang, X. (2017) Engineering CAR-T cells. *Biomark Res*, **5**, 22.
283. Turtle, C.J., Hudecek, M., Jensen, M.C. and Riddell, S.R. (2012) Engineered T cells for anti-cancer therapy. *Curr Opin Immunol*, **24**, 633-639.
284. Benmebarek, M.R., Karches, C.H., Cadilha, B.L., Lesch, S., Endres, S. and Kobold, S. (2019) Killing Mechanisms of Chimeric Antigen Receptor (CAR) T Cells. *Int J Mol Sci*, **20**.
285. Zhao, L. and Cao, Y.J. (2019) Engineered T Cell Therapy for Cancer in the Clinic. *Front Immunol*, **10**, 2250.
286. Panjwani, M.K., Smith, J.B., Schutsky, K., Gnanandarajah, J., O'Connor, C.M., Powell, D.J., Jr. and Mason, N.J. (2016) Feasibility and Safety of RNA-transfected CD20-specific Chimeric Antigen Receptor T Cells in Dogs with Spontaneous B Cell Lymphoma. *Mol Ther*, **24**, 1602-1614.
287. Zhao, Y., Moon, E., Carpenito, C., Paulos, C.M., Liu, X., Brennan, A.L., Chew, A., Carroll, R.G., Scholler, J., Levine, B.L. *et al.* (2010) Multiple injections of electroporated autologous T cells expressing a chimeric antigen receptor mediate regression of human disseminated tumor. *Cancer Res*, **70**, 9053-9061.
288. McGarrity, G.J., Hoyah, G., Winemiller, A., Andre, K., Stein, D., Blick, G., Greenberg, R.N., Kinder, C., Zolopa, A., Binder-Scholl, G. *et al.* (2013) Patient monitoring and follow-up in lentiviral clinical trials. *J Gene Med*, **15**, 78-82.
289. Wang, X. and Riviere, I. (2016) Clinical manufacturing of CAR T cells: foundation of a promising therapy. *Mol Ther Oncolytics*, **3**, 16015.
290. Foster, J.B., Choudhari, N., Perazzelli, J., Storm, J., Hofmann, T.J., Jain, P., Storm, P.B., Pardi, N., Weissman, D., Waanders, A.J. *et al.* (2019) Purification of mRNA Encoding Chimeric Antigen Receptor Is Critical for Generation of a Robust T-Cell Response. *Hum Gene Ther*, **30**, 168-178.
291. Beatty, G.L., Haas, A.R., Maus, M.V., Torigian, D.A., Soulen, M.C., Plesa, G., Chew, A., Zhao, Y., Levine, B.L., Albelda, S.M. *et al.* (2014) Mesothelin-specific chimeric antigen receptor mRNA-engineered T cells induce anti-tumor activity in solid malignancies. *Cancer Immunol Res*, **2**, 112-120.
292. Barrett, D.M., Zhao, Y., Liu, X., Jiang, S., Carpenito, C., Kalos, M., Carroll, R.G., June, C.H. and Grupp, S.A. (2011) Treatment of advanced leukemia in mice with mRNA engineered T cells. *Hum Gene Ther*, **22**, 1575-1586.
293. Iriguchi, S. and Kaneko, S. (2019) Toward the development of true "off-the-shelf" synthetic T-cell immunotherapy. *Cancer Sci*, **110**, 16-22.
294. Nianias, A. and Themeli, M. (2019) Induced Pluripotent Stem Cell (iPSC)-Derived Lymphocytes for Adoptive Cell Immunotherapy: Recent Advances and Challenges. *Curr Hematol Malig Rep*, **14**, 261-268.
295. Imai, C., Mihara, K., Andreansky, M., Nicholson, I.C., Pui, C.H., Geiger, T.L. and Campana, D. (2004) Chimeric receptors with 4-1BB signaling capacity provoke potent cytotoxicity against acute lymphoblastic leukemia. *Leukemia*, **18**, 676-684.
296. Stancheva, I. (2011) Revisiting heterochromatin in embryonic stem cells. *PLoS Genet*, **7**, e1002093.
297. Golob, J.L., Paige, S.L., Muskheli, V., Pabon, L. and Murry, C.E. (2008) Chromatin remodeling during mouse and human embryonic stem cell differentiation. *Dev Dyn*, **237**, 1389-1398.
298. Norrman, K., Fischer, Y., Bonnamy, B., Wolfhagen Sand, F., Ravassard, P. and Semb, H. (2010) Quantitative comparison of constitutive promoters in human ES cells. *PloS one*, **5**, e12413.

299. Wang, J., Song, L., Grover, D., Azrak, S., Batzer, M.A. and Liang, P. (2006) dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum Mutat*, **27**, 323-329.
300. Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. and Gerstein, M.B. (2010) Annotating non-coding regions of the genome. *Nat Rev Genet*, **11**, 559-571.
301. Penzkofer, T., Jager, M., Figlerowicz, M., Badge, R., Mundlos, S., Robinson, P.N. and Zemojtel, T. (2017) L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res*, **45**, D68-D73.
302. Penzkofer, T., Dandekar, T. and Zemojtel, T. (2005) L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res*, **33**, D498-500.
303. Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M. and Moran, J.V. (2010) LINE-1 retrotransposition activity in human genomes. *Cell*, **141**, 1159-1170.
304. Antonarakis, S.E. (1995) Molecular genetics of coagulation factor VIII gene and hemophilia A. *Thromb Haemost*, **74**, 322-328.
305. Matrai, J., Chuah, M.K. and VandenDriessche, T. (2010) Preclinical and clinical progress in hemophilia gene therapy. *Curr Opin Hematol*, **17**, 387-392.
306. Garber, K. (2000) rFactor VIII deficit questioned. *Nat Biotechnol*, **18**, 1133.
307. Doering, C.B. and Spencer, H.T. (2009) Advancements in gene transfer-based therapy for hemophilia A. *Expert Rev Hematol*, **2**, 673-683.
308. Mannully, S.T., L, N.R. and Pulicherla, K.K. (2018) Perspectives on progressive strategies and recent trends in the production of recombinant human factor VIII. *Int J Biol Macromol*, **119**, 496-504.
309. Santagostino, E. (2014) A new recombinant factor VIII: from genetics to clinical use. *Drug Des Devel Ther*, **8**, 2507-2515.
310. Thompson, A.R. (2003) Structure and function of the factor VIII gene and protein. *Semin Thromb Hemost*, **29**, 11-22.
311. Bowen, D.J. (2002) Haemophilia A and haemophilia B: molecular insights. *Mol Pathol*, **55**, 127-144.
312. Jennifer Johnston, C.B.D.a.H.T.S. (2011) *Gene Therapy Strategies Incorporating Large Transgenes*, Gene Therapy - Developments and Future Perspectives.
313. Arruda, V.R. (2006) Toward gene therapy for hemophilia A with novel adenoviral vectors: successes and limitations in canine models. *J Thromb Haemost*, **4**, 1215-1217.
314. Brunetti-Pierri, N., Palmer, D.J., Beaudet, A.L., Carey, K.D., Finegold, M. and Ng, P. (2004) Acute toxicity after high-dose systemic injection of helper-dependent adenoviral vectors into nonhuman primates. *Hum Gene Ther*, **15**, 35-46.
315. Wu, Z., Yang, H. and Colosi, P. (2010) Effect of genome size on AAV vector packaging. *Mol Ther*, **18**, 80-86.
316. Lu, H., Chen, L., Wang, J., Huack, B., Sarkar, R., Zhou, S., Xu, R., Ding, Q., Wang, X., Wang, H. *et al.* (2008) Complete correction of hemophilia A with adeno-associated viral vectors containing a full-size expression cassette. *Hum Gene Ther*, **19**, 648-654.
317. Kumar, M., Keller, B., Makalou, N. and Sutton, R.E. (2001) Systematic determination of the packaging limit of lentiviral vectors. *Hum Gene Ther*, **12**, 1893-1905.
318. al Yacoub, N., Romanowska, M., Haritonova, N. and Foerster, J. (2007) Optimized production and concentration of lentiviral vectors containing large inserts. *J Gene Med*, **9**, 579-584.
319. Radcliffe, P.A., Sion, C.J., Wilkes, F.J., Custard, E.J., Beard, G.L., Kingsman, S.M. and Mitrophanous, K.A. (2008) Analysis of factor VIII mediated suppression of lentiviral vector titres. *Gene Ther*, **15**, 289-297.
320. Pipe, S.W. (2009) Functional roles of the factor VIII B domain. *Haemophilia*, **15**, 1187-1196.
321. Ohlfest, J.R., Frandsen, J.L., Fritz, S., Lobitz, P.D., Perkinson, S.G., Clark, K.J., Nelsestuen, G., Key, N.S., McIvor, R.S., Hackett, P.B. *et al.* (2005) Phenotypic correction and long-term expression of factor VIII in hemophilic mice by immunotolerization and nonviral gene transfer using the Sleeping Beauty transposon system. *Blood*, **105**, 2691-2698.
322. Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y. and Xu, T. (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, **122**, 473-483.
323. Matsui, H., Fujimoto, N., Sasakawa, N., Ohinata, Y., Shima, M., Yamanaka, S., Sugimoto, M. and Hotta, A. (2014) Delivery of full-length factor VIII using a piggyBac transposon vector to correct a mouse model of hemophilia A. *PLoS One*, **9**, e104957.
324. Essner, J.J., McIvor, R.S. and Hackett, P.B. (2005) Awakening gene therapy with Sleeping Beauty transposons. *Curr Opin Pharmacol*, **5**, 513-519.

325. Ghirlando, R. and Felsenfeld, G. (2016) CTCF: making the right connections. *Genes Dev*, **30**, 881-891.
326. Ong, C.T. and Corces, V.G. (2014) CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, **15**, 234-246.
327. Arzate-Mejia, R.G., Recillas-Targa, F. and Corces, V.G. (2018) Developing in 3D: the role of CTCF in cell differentiation. *Development*, **145**.
328. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381-385.
329. Cremer, T. and Cremer, M. (2010) Chromosome territories. *Cold Spring Harb Perspect Biol*, **2**, a003889.
330. Merkenschlager, M. and Nora, E.P. (2016) CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet*, **17**, 17-43.
331. Hansen, A.S., Cattoglio, C., Darzacq, X. and Tjian, R. (2018) Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*, **9**, 20-32.
332. Skibbens, R.V. (2016) Of Rings and Rods: Regulating Cohesin Entrapment of DNA to Generate Intra- and Intermolecular Tethers. *PLoS Genet*, **12**, e1006337.
333. Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N. and Peters, J.M. (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, **544**, 503-507.
334. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L.A. (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*, **15**, 2038-2049.
335. Alipour, E. and Marko, J.F. (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res*, **40**, 11202-11212.
336. Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenko, V.V. and Ren, B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, **128**, 1231-1245.
337. Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H. and Dekker, C. (2018) Real-time imaging of DNA loop extrusion by condensin. *Science*, **360**, 102-105.
338. Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R. and Darzacq, X. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife*, **6**.
339. Kim, H.Y. (2006) Improved expression vector activity using insulators and scaffold/matrix-attachment regions. *BioProcess International*.
340. Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L. and Caiafa, P. (2008) CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem*, **283**, 21873-21880.
341. Morshedi, A., Soroush Noghabi, M. and Droge, P. (2013) Use of UTF1 genetic control elements as iPSC reporter. *Stem Cell Rev Rep*, **9**, 523-530.
342. Tan, S.M., Wang, S.T., Hentze, H. and Droge, P. (2007) A UTF1-based selection system for stable homogeneously pluripotent human embryonic stem cell cultures. *Nucleic Acids Res*, **35**, e118.
343. Lavie, L., Maldener, E., Brouha, B., Meese, E.U. and Mayer, J. (2004) The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res*, **14**, 2253-2260.
344. Li, J., Kannan, M., Trivett, A.L., Liao, H., Wu, X., Akagi, K. and Symer, D.E. (2014) An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. *Nucleic Acids Res*, **42**, 4546-4562.
345. Schumann, G.G., Fuchs, N.V., Tristan-Ramos, P., Sebe, A., Ivics, Z. and Heras, S.R. (2019) The impact of transposable element activity on therapeutically relevant human stem cells. *Mob DNA*, **10**, 9.
346. Minakami, R., Kurose, K., Etoh, K., Furuhashi, Y., Hattori, M. and Sakaki, Y. (1992) Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. *Nucleic Acids Res*, **20**, 3139-3145.
347. Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol*, **10**, 6718-6729.
348. Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol*, **21**, 1973-1985.

349. Denli, A.M., Narvaiza, I., Kerman, B.E., Pena, M., Benner, C., Marchetto, M.C., Diedrich, J.K., Aslanian, A., Ma, J., Moresco, J.J. *et al.* (2015) Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell*, **163**, 583-593.
350. Babatz, T.D. and Burns, K.H. (2013) Functional impact of the human mobilome. *Curr Opin Genet Dev*, **23**, 264-270.
351. Burns, K.H. and Boeke, J.D. (2012) Human transposon tectonics. *Cell*, **149**, 740-752.
352. Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S. and Moran, J.V. (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet*, **16**, 1569-1577.
353. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian, H.H., Jr. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917-927.
354. Faulkner, G.J. and Billon, V. (2018) L1 retrotransposition in the soma: a field jumping ahead. *Mob DNA*, **9**, 22.
355. Bodak, M., Yu, J. and Ciaudo, C. (2014) Regulation of LINE-1 in mammals. *Biomol Concepts*, **5**, 409-428.
356. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. and Kazazian, H.H., Jr. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A*, **100**, 5280-5285.
357. Ergun, S., Buschmann, C., Heukeshoven, J., Dammann, K., Schnieders, F., Lauke, H., Chalajour, F., Kilic, N., Stratling, W.H. and Schumann, G.G. (2004) Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J Biol Chem*, **279**, 27753-27763.
358. Belgnaoui, S.M., Gosden, R.G., Semmes, O.J. and Haoudi, A. (2006) Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int*, **6**, 13.
359. Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X. and Ramalho-Santos, M. (2018) A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*, **174**, 391-405 e319.
360. Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S. and Devine, S.E. (2006) Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet*, **78**, 671-679.
361. Chen, X. and Goncalves, M.A. (2016) Engineered Viruses as Genome Editing Devices. *Mol Ther*, **24**, 447-457.
362. Hardee, C.L., Arevalo-Soliz, L.M., Hornstein, B.D. and Zechiedrich, L. (2017) Advances in Non-Viral DNA Vectors for Gene Therapy. *Genes (Basel)*, **8**.
363. Provasi, E., Genovese, P., Lombardo, A., Magnani, Z., Liu, P.Q., Reik, A., Chu, V., Paschon, D.E., Zhang, L., Kuball, J. *et al.* (2012) Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nat Med*, **18**, 807-815.
364. Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordonez, A., Hannan, N.R., Rouhani, F.J. *et al.* (2011) Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, **478**, 391-394.
365. Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F. *et al.* (2011) In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*, **29**, 1717-1726.
366. Soldner, F., Laganieri, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, **146**, 318-331.
367. Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol*, **26**, 808-816.
368. Yu, W., Mookherjee, S., Chaitankar, V., Hiriyanna, S., Kim, J.W., Brooks, M., Ataeijannati, Y., Sun, X., Dong, L., Li, T. *et al.* (2017) Nrl knockdown by AAV-delivered CRISPR/Cas9 prevents retinal degeneration in mice. *Nat Commun*, **8**, 14716.
369. Young, C.S., Hicks, M.R., Ermolova, N.V., Nakano, H., Jan, M., Younesi, S., Karumbayaram, S., Kumagai-Cresse, C., Wang, D., Zack, J.A. *et al.* (2016) A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells. *Cell Stem Cell*, **18**, 533-540.

370. Ren, J., Zhang, X., Liu, X., Fang, C., Jiang, S., June, C.H. and Zhao, Y. (2017) A versatile system for rapid multiplex genome-edited CAR T cell generation. *Oncotarget*, **8**, 17002-17011.
371. Khosravi, M.A., Abbasalipour, M., Concordet, J.P., Berg, J.V., Zeinali, S., Arashkia, A., Azadmanesh, K., Buch, T. and Karimipour, M. (2019) Targeted deletion of BCL11A gene by CRISPR-Cas9 system for fetal hemoglobin reactivation: A promising approach for gene therapy of beta thalassemia disease. *Eur J Pharmacol*, **854**, 398-405.
372. Sanches-da-Silva, G.N., Medeiros, L.F.S. and Lima, F.M. (2019) The Potential Use of the CRISPR-Cas System for HIV-1 Gene Therapy. *Int J Genomics*, **2019**, 8458263.
373. Schwank, G., Koo, B.K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K. *et al.* (2013) Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*, **13**, 653-658.
374. Gupta, R.M. and Musunuru, K. (2014) Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest*, **124**, 4154-4161.
375. Azaro M.A. and Landy, A. (2002) λ Int and the λ Int family. *Mobile DNA II. ASM Press, Washington, DC*, In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A. (eds), 118-148.
376. Gardner, J.F. and Nash, H.A. (1986) Role of Escherichia coli IHF protein in lambda site-specific recombination. A mutational analysis of binding sites. *J Mol Biol*, **191**, 181-189.
377. Richet, E., Abcarian, P. and Nash, H.A. (1986) The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell*, **46**, 1011-1021.
378. Lange-Gustafson, B.J. and Nash, H.A. (1984) Purification and properties of Int-h, a variant protein involved in site-specific recombination of bacteriophage lambda. *J Biol Chem*, **259**, 12724-12732.
379. Tan, S.M. and Droge, P. (2005) Comparative analysis of sequence-specific DNA recombination systems in human embryonic stem cells. *Stem Cells*, **23**, 868-873.
380. Kotewicz, M., Chung, S., Takeda, Y. and Echols, H. (1977) Characterization of the integration protein of bacteriophage lambda as a site-specific DNA-binding protein. *Proc Natl Acad Sci U S A*, **74**, 1511-1515.
381. Sadelain, M., Riviere, I. and Riddell, S. (2017) Therapeutic T cell engineering. *Nature*, **545**, 423-431.
382. Themeli, M., Kloss, C.C., Ciriello, G., Fedorov, V.D., Perna, F., Gonen, M. and Sadelain, M. (2013) Generation of tumor-targeted human T lymphocytes from induced pluripotent stem cells for cancer therapy. *Nature biotechnology*, **31**, 928-933.
383. Pezic, D., Manakov, S.A., Sachidanandam, R. and Aravin, A.A. (2014) piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev*, **28**, 1410-1428.
384. Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S.M., Turelli, P. and Trono, D. (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev*, **28**, 1397-1409.
385. Ishiuchi, T. and Torres-Padilla, M.E. (2014) LINEing germ and embryonic stem cells' silencing of retrotransposons. *Genes Dev*, **28**, 1381-1383.
386. Singh, H., Figliola, M.J., Dawson, M.J., Olivares, S., Zhang, L., Yang, G., Maiti, S., Manuri, P., Senyukov, V., Jena, B. *et al.* (2013) Manufacture of clinical-grade CD19-specific T cells stably expressing chimeric antigen receptor using Sleeping Beauty system and artificial antigen presenting cells. *PLoS One*, **8**, e64138.
387. Peng, P.D., Cohen, C.J., Yang, S., Hsu, C., Jones, S., Zhao, Y., Zheng, Z., Rosenberg, S.A. and Morgan, R.A. (2009) Efficient nonviral Sleeping Beauty transposon-based TCR gene transfer to peripheral blood lymphocytes confers antigen-specific antitumor reactivity. *Gene Ther*, **16**, 1042-1049.
388. Sather, B.D., Romano Ibarra, G.S., Sommer, K., Curinga, G., Hale, M., Khan, I.F., Singh, S., Song, Y., Gwiazda, K., Sahni, J. *et al.* (2015) Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci Transl Med*, **7**, 307ra156.
389. Schumann, K., Lin, S., Boyer, E., Simeonov, D.R., Subramaniam, M., Gate, R.E., Haliburton, G.E., Ye, C.J., Bluestone, J.A., Doudna, J.A. *et al.* (2015) Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc Natl Acad Sci U S A*, **112**, 10437-10442.
390. Roth, T.L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P.J., Hiatt, J., Saco, J., Krystofinski, P., Li, H. *et al.* (2018) Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature*, **559**, 405-409.

391. Liu, J., Zhou, G., Zhang, L. and Zhao, Q. (2019) Building Potent Chimeric Antigen Receptor T Cells With CRISPR Genome Editing. *Front Immunol*, **10**, 456.
392. Peng, R., Lin, G. and Li, J. (2016) Potential pitfalls of CRISPR/Cas9-mediated genome editing. *FEBS J*, **283**, 1218-1231.
393. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. and Taipale, J. (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med*, **24**, 927-930.
394. Ren, J., Liu, X., Fang, C., Jiang, S., June, C.H. and Zhao, Y. (2017) Multiplex Genome Editing to Generate Universal CAR T Cells Resistant to PD1 Inhibition. *Clin Cancer Res*, **23**, 2255-2266.
395. Schaefer, K.A., Wu, W.H., Colgan, D.F., Tsang, S.H., Bassuk, A.G. and Mahajan, V.B. (2017) Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nat Methods*, **14**, 547-548.
396. Christ, N., Corona, T. and Droge, P. (2002) Site-specific recombination in eukaryotic cells mediated by mutant lambda integrases: implications for synaptic complex formation and the reactivity of episomal DNA segments. *Journal of molecular biology*, **319**, 305-314.
397. Muller-Hill, B. (2006) What is life? The paradigm of DNA and protein cooperation at high local concentrations. *Mol Microbiol*, **60**, 253-255.
398. Peter Dröge, B.M.H. (2001) High local protein concentrations at promoters: Strategies in prokaryotic and eukaryotic cells. *BioEssays*, **23**, 179-183.
399. Oehler, S. and Muller-Hill, B. (2010) High local concentration: a fundamental strategy of life. *J Mol Biol*, **395**, 242-253.
400. Spencer, H.T., Denning, G., Gautney, R.E., Dropulic, B., Roy, A.J., Baranyi, L., Gangadharan, B., Parker, E.T., Lollar, P. and Doering, C.B. (2011) Lentiviral vector platform for production of bioengineered recombinant coagulation factor VIII. *Mol Ther*, **19**, 302-309.
401. Fontes, A.M., Melo, F.U., Greene, L.J., Faca, V.M., Lin, Y., Gerson, S.L. and Covas, D.T. (2012) Production of human factor VIII-FL in 293T cells using the bicistronic MGMT(P140K)-retroviral vector. *Genet Mol Res*, **11**, 775-789.
402. Fu, Y., Sinha, M., Peterson, C.L. and Weng, Z. (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet*, **4**, e1000138.
403. Kim, S., Yu, N.K. and Kaang, B.K. (2015) CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med*, **47**, e166.
404. Zlotorynski, E. (2015) Chromosome biology: CTCF-binding site orientation shapes the genome. *Nat Rev Mol Cell Biol*, **16**, 578-579.
405. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y. *et al.* (2015) CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, **162**, 900-910.
406. Devon Zimmerman, K.P., Matthew Hall, Jacob Elmer. (2018) Enhancement of Transgene Expression by NF-Y and CTCF. *BioRxiv*.
407. Ardeljan, D., Taylor, M.S., Ting, D.T. and Burns, K.H. (2017) The Human Long Interspersed Element-1 Retrotransposon: An Emerging Biomarker of Neoplasia. *Clin Chem*, **63**, 816-822.

APPENDIX

Table A1: List of genomic coordinates of the targeted *attH4X* in HT1080 bulk culture along with their frequency of occurrence obtained using deep paired-end sequencing [reprinted with permission (181)].

Chromosome	Start	End	Counts
10	47432566	47432842	14116
5	37655668	37655966	3562
11	86542419	86542586	3223
1	210886727	210886826	312
X	104143467	104143534	234
2	224486415	224486548	125
3	137376024	137376055	115
1	107845629	107845731	94
1	205900757	205900858	88
X	132864468	132864513	80
9	15050823	15050879	66
12	33493949	33494006	65
3	67541275	67541444	62
3	156067893	156067992	57
12	102493125	102493166	56
3	96092506	96092604	52
2	58169733	58169815	50
3	169414716	169414795	50
7	17353179	17353266	50
18	7712701	7712801	45
3	151886483	151886577	42
Y	19661244	19661283	41
10	109547670	109547704	40
11	3269140	3269321	40
GL000213.1	22304	22340	39
7	133969091	133969162	36
14	22687362	22687421	33
11	63394044	63394123	32
5	157722276	157722356	32
12	97866922	97866956	31
7	94821074	94821154	31
1	157346999	157347052	29
1	104058375	104058472	27
5	6648494	6648592	26
1	65742930	65742968	23
5	90032846	90032926	23
3	87652598	87652653	20
5	72981677	72981777	20
7	94968892	94968972	20

1	74015741	74015800	19
13	97594574	97594642	19
13	48395563	48395662	18
12	80082999	80083088	17
15	49006499	49006529	17
1	169409192	169409301	15
13	19653740	19653819	15
3	66894359	66894441	15
3	164162167	164162247	15
7	7392706	7392806	15
X	56195456	56195495	15
X	57272103	57272162	15
1	104587615	104587780	14
4	120510442	120510519	14
X	131657562	131657608	14
X	139064922	139065002	14
6	101063329	101063409	13
7	18011687	18011843	13
7	97052219	97052319	13
13	112815802	112815927	12
2	159823285	159823360	12
21	46631673	46631755	12
X	67831307	67831343	12
12	63338237	63338266	11
2	112613731	112613770	11
1	230475381	230475453	10
12	88438302	88438404	10
13	28782949	28783027	10
13	42451916	42451996	10
4	55168854	55168933	10
6	70805252	70805286	10
1	210963630	210963703	9
11	49964843	49964882	9
14	54406590	54406670	9
17	35525040	35525120	9
18	8150379	8150477	9
2	48301703	48301783	9
5	162328591	162328667	9
7	121292246	121292348	9
15	76097854	76097898	8
4	124013145	124013175	8
4	160341227	160341303	8
8	110255556	110255618	8
8	119378032	119378094	8

1	100821321	100821351	7
12	101061615	101061695	7
4	129054501	129054551	7
7	21928644	21928715	7
KI270728.1	1154641	1154677	7
X	76381100	76381131	7
X	117466937	117466996	7
1	174710046	174710098	6
10	91645492	91645572	6
3	133735635	133735717	6
4	127076083	127076145	6
6	106406212	106406289	6
8	65381296	65381361	6
X	21546258	21546334	6
X	43726512	43726618	6
X	57745517	57745556	6
X	112182436	112182578	6
Y	14270975	14271004	6
Y	18957360	18957395	6
11	110849041	110849087	5
11	110849113	110849266	5
18	44365737	44365816	5
2	175913051	175913206	5
2	212056109	212056189	5
21	14622503	14622533	5
4	20721077	20721108	5
4	174619360	174619541	5
5	126087347	126087376	5
8	69281522	69281599	5
8	106562303	106562382	5
9	78055601	78055630	5
10	75577221	75577300	4
12	90754213	90754277	4
12	94065653	94065705	4
12	112157957	112157999	4
14	37489209	37489378	4
18	70238091	70238130	4
2	101458962	101458992	4
3	3890347	3890418	4
3	66311241	66311295	4
3	121263557	121263723	4
4	100867544	100867574	4
6	122987530	122987610	4
7	11390351	11390423	4

8	48168112	48168142	4
9	104704788	104704871	4
X	39543352	39543502	4
X	54846100	54846213	4
X	65893404	65893434	4
X	93662021	93662101	4
1	92254099	92254179	3
1	173030869	173030919	3
10	64809195	64809267	3
10	108695078	108695158	3
10	109338348	109338377	3
11	48452541	48452580	3
12	20335360	20335429	3
13	42775696	42775751	3
13	49881048	49881124	3
13	112815595	112815726	3
19	32786675	32786705	3
2	62621158	62621289	3
2	72299236	72299271	3
2	189891984	189892064	3
20	18940603	18940682	3
3	14193528	14193567	3
3	29984162	29984238	3
3	132745905	132745935	3
3	132958085	132958155	3
3	151738746	151738785	3
3	161708398	161708480	3
3	189257757	189257833	3
4	2921883	2921961	3
4	52091040	52091227	3
4	129946708	129946748	3
4	138784371	138784406	3
5	55097912	55097941	3
5	72807787	72807867	3
5	130608117	130608199	3
6	64569326	64569404	3
6	129836807	129836836	3
7	42370174	42370204	3
7	126278796	126278826	3
8	70464804	70464834	3
9	77266893	77266924	3
9	86089508	86089570	3
9	106203067	106203115	3
9	135842204	135842260	3

X	25883671	25883751	3
X	66859096	66859158	3
X	72890882	72890961	3
X	76266404	76266576	3
X	77822728	77822791	3
X	114184815	114184894	3
X	123284616	123284662	3
Y	17244190	17244225	3
1	100201264	100201315	2
1	113574938	113575015	2
1	221048769	221048848	2
10	27588649	27588725	2
10	74607983	74608053	2
10	85279118	85279180	2
11	10964600	10964629	2
11	13991414	13991525	2
11	23485409	23485442	2
12	298882	298961	2
12	77287529	77287608	2
12	100303003	100303060	2
14	37189162	37189192	2
14	37631864	37631943	2
14	38600553	38600628	2
14	60528004	60528034	2
15	97584199	97584253	2
2	55836250	55836302	2
2	108430772	108430852	2
2	135517023	135517103	2
2	185687111	185687141	2
2	209242803	209242841	2
2	210985966	210986002	2
2	215989728	215989806	2
3	1062530	1062583	2
3	37430107	37430137	2
3	37659431	37659483	2
3	66293038	66293091	2
3	83107843	83107878	2
3	118321439	118321542	2
3	137318338	137318505	2
3	181003254	181003292	2
4	141164135	141164215	2
4	151596775	151596822	2
5	10179015	10179090	2
5	23187510	23187541	2

5	58607294	58607324	2
5	62036387	62036469	2
5	63091672	63091711	2
5	89801765	89801794	2
5	131897639	131897669	2
5	157425693	157425722	2
6	44880299	44880339	2
6	74034428	74034458	2
6	80684596	80684626	2
6	85203246	85203406	2
6	103523273	103523307	2
6	118839175	118839250	2
6	145674670	145674700	2
7	82491495	82491547	2
7	83628538	83628668	2
7	90550302	90550332	2
7	103404937	103405067	2
8	24163281	24163355	2
8	50991236	50991272	2
8	126458501	126458603	2
9	4468020	4468049	2
9	13793024	13793059	2
X	22233561	22233601	2
X	37529185	37529262	2
X	39279450	39279485	2
X	65065894	65065929	2
X	68777046	68777075	2
X	72829874	72829954	2
X	73893754	73893927	2
X	78231808	78231854	2
X	132450416	132450455	2
Y	19797122	19797152	2
Y	25135576	25135605	2
1	49653699	49653732	1
1	95798603	95798636	1
1	104058227	104058328	1
1	121585910	121585971	1
1	183683918	183683974	1
1	217381027	217381057	1
10	25620855	25620916	1
10	32501130	32501208	1
10	38188104	38188231	1
10	43835326	43835400	1
10	95933132	95933212	1

11	2497417	2497464	1
11	10340974	10341023	1
11	10951117	10951204	1
11	18993428	18993458	1
11	21696140	21696178	1
11	35023350	35023379	1
11	40849880	40849915	1
11	60010348	60010383	1
11	102818644	102818711	1
11	103911692	103911792	1
11	106757099	106757146	1
12	67554440	67554498	1
12	83682922	83682956	1
12	87419901	87419977	1
12	88668816	88668865	1
13	28692799	28692837	1
13	48138356	48138391	1
13	48417665	48417789	1
13	48674351	48674392	1
13	65108638	65108684	1
13	82954030	82954059	1
13	110078372	110078512	1
13	112815740	112815792	1
14	23943808	23943842	1
14	52595569	52595599	1
14	57099859	57100019	1
14	86047199	86047299	1
15	38802218	38802296	1
16	47773943	47773979	1
17	60765937	60765967	1
18	7867947	7868076	1
18	70922054	70922084	1
18	79082158	79082309	1
19	20989556	20989595	1
19	34566964	34567007	1
2	24536202	24536232	1
2	37940636	37940717	1
2	114303065	114303136	1
2	115893829	115893883	1
2	115964762	115964810	1
2	138797304	138797352	1
2	157705827	157705877	1
2	160627322	160627351	1
2	175710458	175710523	1

2	179630901	179630930	1
2	181749287	181749327	1
2	195952731	195952831	1
2	197749565	197749643	1
2	198686171	198686206	1
20	5240493	5240528	1
20	7319855	7319984	1
20	13053739	13053800	1
20	17412412	17412441	1
21	32220857	32220891	1
3	1654266	1654320	1
3	4244989	4245018	1
3	19681438	19681508	1
3	25932828	25932925	1
3	26322798	26322827	1
3	58677350	58677405	1
3	61262568	61262723	1
3	79049279	79049308	1
3	101923796	101923853	1
3	110496407	110496461	1
3	119131435	119131463	1
3	131043495	131043524	1
3	132607456	132607536	1
3	162248018	162248047	1
3	173383511	173383568	1
4	36438850	36438880	1
4	52152806	52152868	1
4	54826081	54826116	1
4	62570692	62570725	1
4	63590979	63591014	1
4	88540328	88540357	1
4	89089117	89089159	1
4	93931541	93931585	1
4	105298432	105298463	1
4	105410291	105410324	1
4	106470321	106470368	1
4	127462897	127462977	1
4	134441891	134441921	1
4	149340304	149340484	1
4	158901718	158901771	1
4	161116977	161117008	1
4	180469748	180469818	1
4	188673825	188673905	1
5	11671624	11671722	1

5	12154880	12154930	1
5	12673020	12673064	1
5	34446538	34446577	1
5	36466573	36466648	1
5	43964701	43964750	1
5	64475460	64475491	1
5	111715962	111716031	1
5	156784195	156784224	1
5	171021523	171021680	1
6	4406975	4407041	1
6	6071466	6071497	1
6	6669546	6669575	1
6	23259034	23259078	1
6	29361162	29361192	1
6	54328854	54328905	1
6	56205318	56205396	1
6	57289186	57289215	1
6	61184838	61184882	1
6	62087944	62087999	1
6	64897837	64897887	1
6	76098911	76098978	1
6	80241002	80241031	1
6	80432412	80432451	1
6	84822318	84822357	1
6	85877646	85877676	1
6	85891427	85891480	1
6	93835946	93836020	1
6	115479811	115479840	1
6	117646938	117646991	1
6	118422412	118422441	1
6	120815032	120815057	1
6	127412009	127412038	1
6	130172958	130173105	1
6	143034121	143034146	1
6	143856165	143856214	1
6	144507015	144507087	1
6	147252460	147252491	1
6	161177395	161177424	1
7	9168104	9168133	1
7	24353484	24353539	1
7	31928150	31928200	1
7	40933685	40933715	1
7	76993700	76993783	1
7	81479819	81479849	1

7	111704966	111705023	1
7	123855950	123856021	1
8	9817216	9817350	1
8	59382190	59382229	1
8	60449422	60449474	1
8	62977446	62977510	1
8	94126185	94126228	1
8	112875671	112875814	1
9	1394768	1394814	1
9	19766048	19766097	1
9	21650599	21650633	1
9	27784005	27784102	1
9	75326431	75326480	1
9	100618372	100618401	1
9	114871042	114871112	1
X	3983375	3983414	1
X	6974962	6975007	1
X	8936147	8936182	1
X	17573430	17573465	1
X	55094658	55094689	1
X	55605269	55605348	1
X	55673094	55673123	1
X	56675050	56675084	1
X	56810500	56810557	1
X	57480993	57481029	1
X	64476702	64476737	1
X	64660959	64660990	1
X	78373062	78373133	1
X	81065026	81065055	1
X	96428324	96428357	1
X	108943152	108943212	1
X	110658261	110658296	1
X	110972904	110972974	1
X	113734752	113734813	1
X	115984351	115984421	1
X	122062343	122062461	1
X	148180774	148180871	1
X	155320936	155320974	1
Y	3594102	3594143	1
Y	15618152	15618191	1
Y	15853088	15853128	1
Y	20675850	20675879	1
Y	22498980	22499009	1

PUBLICATIONS

1. Harshyaa Makhija, **Suki Roy**, Shawn Hoon, Farid John Ghadessy, Desmond Wong, Rahul Jaiswal, Dario Campana and Peter Dröge (2018) A novel λ integrase-mediated seamless vector transgenesis platform for therapeutic protein expression. Nucleic Acids Research 46(16), e99. DOI: 10.1093/nar/gky500
2. Namrata Chaudhari, Amanda Rickard, **Suki Roy**, Peter Dröge, Harshyaa Makhija (2020) A versatile transgenesis platform for safe and functional delivery of large seamless DNA cargo. Stem Cell Research and Therapy (in press).

CONFERENCE PRESENTATIONS

Poster Presentation

1. **Science Improving Lives Symposium (SIL)** organized by Nanyang Technological University on 27th February 2017 at Nanyang Technological University, Singapore.
Title: Site-specific seamless vector transgenesis in human *LINE-1 elements* for therapeutic protein expression
2. **International Congress of Cell Biology (ICCB)** organized by CSIR-Centre for Cellular and Molecular Biology on 27th - 31st January 2018 at Hyderabad, India.
Title: Site-specific seamless vector transgenesis in human *LINE-1 elements*
3. **Cambridge International Stem Cell Symposium** organized by Wellcome - MRC Cambridge Stem Cell Institute on 19th – 21st September 2018 at Cambridge, United Kingdom
Title: Site-specific seamless vector transgenesis in human *LINE-1 elements*
4. **27th Annual Congress** organized by the European Society of Gene and Cell Therapy (ESGCT) on 22nd - 25th October 2019 at Barcelona, Spain.
Title: A novel λ integrase-mediated transgenesis platform for biologics production

Oral Presentations

1. **3 Minute Thesis NTU (3MT)** on 20th July 2018 in Singapore
Title: DNA Scissors
2. **3 Minute Thesis Singapore (3MT)** on 27th July 2018 in Singapore
Title: DNA Scissors
3. **Falling Walls Lab Singapore** on 5th September 2018 in Singapore
Title: Breaking the Wall of Human Gene Therapy
4. **Young SBS Scientist Seminar Series** on 18th September 2019 in Singapore
Title: A novel λ integrase mediated transgenesis platform for therapeutic protein expression