DEVELOPING THERAPEUTIC BIOIMPLANTS FOR HEMOPHILIA A: BIOSAFETY EVALUATION OF TARGETED TRANSGENE INTEGRATION IN PRIMARY HUMAN UMBILICAL CORD-LINING CELLS USING PHIC31 INTEGRASE AND ZINC FINGER NUCLEASES

JAICHANDRAN SIVALINGAM

[B.Sc., Murdoch University, Western Australia]

A THESIS SUBMITTED FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

YONG LOO LIN SCHOOL OF MEDICINE

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Jaichandran

8th of July 2013

ACKNOWLEDGEMENTS

"A journey of a thousand miles begins with a single step", so read the caption that triggered a chain of events that led to the work currently covered in this doctoral thesis. Indeed, it was an arduous journey that saw me transit from being a full-time staff at National Cancer Centre of Singapore to a part-time student at NUS, through marriage and onwards towards fatherhood. The amount of time, effort and sacrifice put into completion of this thesis cannot be emphasized enough. It was a path wrought with many challenges and failures, countless trouble-shoots, numerous detours and careful re-planning and execution of experiments. Nevertheless, encouragements came at times of disappointments, grace and favors abound when least expected, meticulous advice and ideas were shared when I had exhausted of mine own, inspirations were drawn from most unlikely sources and friendships and collaborations were forged. Thus, I am indebted to many for help rendered in one way or another that led to completion of this thesis.

My most sincere and heartfelt gratitude to my employer and thesis supervisor, Professor Kon Oi Lian cannot be emphasized enough. If there's any single person I could attribute to successful completion of this thesis, it would have to be the ever so eloquent Prof. Kon. The maestro and grand conductor of this orchestra of gene therapy, she's been in the helm of this project, from conceptualization of ideas to trouble-shooting and execution of experiments to editing this thesis and making it suitable for reading. Much confidence and trust was placed in me where many, including myself, would have doubted and given up. She's encouraged me through countless failures and disappointments, educated me meticulously to mature my scientific knowledge and thinking and had persistently supported me throughout my research career. Of course, behind every successful woman there's an equally brilliant man, Professor Peter Hwang, to whom I am grateful for his kindness, friendship and encouragements.

I am much appreciative of my co-supervisor, Assoc. Professor Phan Toan Thang (Department of Surgery, NUS) for providing valuable samples and reagents necessary for our experiments and for helpful discussions, collaborations and ideas. Much gratitude is owed to collaborators and colleagues who shared their reagents and expertise and offered technical assistance when needed: Dr. Michele P.Calos (Stanford University) for gifts of pTA-attB and pCMV-Int plasmids, Professor Haig Kazazian (University of Pennsylvania) for gifts of FVIII-deficient mice, Professor Boris Fehse (University Medical Centre Hamburg-Eppendorf) for gift of MP71-tCD34-TK007 plasmid, Dr. Mark Richards (Nanyang Polytechnic, Singapore) for gift

TABLE OF CONTENTS

CONTENTS	PAGE

CHAPTER 1

Introduction and literature review

CHAPTER 2

Results and Discussion

cells

CHAPTER 3

General Discussion

CHAPTER 4

Materials and Methods

Appendices

SUMMARY

Gene and cell-based therapies have been shown to successfully treat genetic disorders. However, the inability of current gene therapy vectors to direct transgene integrations precisely into safe genomic sites has been associated with oncogenic mutations and fatal leukemias. A major challenge is devising gene transfer techniques that are efficacious and result in durable transgene expression but are not mutagenic. Using hemophilia A as a disease model, we evaluated the potential of phiC31 integrase and zinc-finger nucleases (ZFNs) to modify primary human umbilical cord-lining epithelial cells (CLECs) to stably integrate and express a factor VIII transgene, with the ultimate aim of developing autologous treatment for pediatric patients with hemophilia A.

PhiC31 integrase-mediated integration of a FVIII transgene cassette into the genome of CLECs achieved durable FVIII secretion for at least 2 months *in vitro*. Retrieved integration events mapped to 18 different pseudo *attP* sites in the genome, with 85% of cells having ≤ 2 copies of the transgene. Transcriptome array analysis of genome-modified cells revealed that 96.5% of genes were unaltered in expression. High-resolution copy number profiling identified 3 genomic regions with minor copy number changes that did not correlate with altered gene expression or integration sites. Spectral karyotyping revealed three different translocations that were rare and nonrecurrent. Integrase-modified cells were not tumorigenic in immunocompromised mice for at least 4 months. Xenoimplantation of FVIII-secreting CLECs in immunocompetent hemophilic mice achieved significant phenotypic correction.

Our phiC31 integrase study revealed that up to 40% of integrations occurred at a particular site within chromosome 8p22. Eight oligoclonal CLECs with transgene integration at the 8p22 site were evaluated for biosafety. A porcine/human hybrid FVIII construct that we assembled showed up to 5-fold increased FVIII expression compared to human FVIII, and was used in this study. Transgenic CLECs secreted high levels of FVIII, did not bear transcriptional profiles of transformed cells, did not have altered genome copy number profiles or gross chromosomal abnormalities, and were not tumorigenic when implanted into immunocompromised mice.

We proceeded to evaluate site-specific ZFNs to promote homologous recombination-mediated integration of transgene cassettes into a proposed safe genomic harbor, the AAVS1 site, in a human cell line and primary human cells. Integration junction PCR and sequencing demonstrated site-specific integration of donor DNAs of up to 9-kb into intron 1 of *PPP1R12C*. A promoter trap strategy integrated donor DNA with splice-acceptor sequences and reduced endogenous

PPP1R12C expression. However, transcriptome data showed no effect on the expression of neighboring genes or potential interacting partners. Targeted integration of hybrid FVIII cassette into the AAVS1 locus induced durable FVIII secretion by CLECs. ZFNs were highly site-specific and only induced low frequency indels at a single intergenic region out of the ten most likely potential off-target sites evaluated by targeted deep sequencing. ZFN-modified CLECs had normal cellular morphology, growth characteristics and chromosomal karyotype. Our results demonstrate the potential of appropriately designed ZFNs as genome-modifying tools for a range of primary human cell types, and merit further development towards the future goal of clinical gene and cell therapy.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aa	amino acid(s)
AAV	adeno-associated virus
ACV	acyclovir
ADA	adenosine deaminase
ADSCs	adipose-derived stromal cells
ALD	adrenoleukodystrophy
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
aPCC	activated prothrombin complex concentrates
ASLV	avian sarcoma-leukosis virus
ATCC	American Type Culture Collection
AVP	arginine vasopressin
AZT	azido-3'-deoxythymidine
BACs	bacterial artificial chromosomes
BDD	B-domain deleted
BGH	bovine growth hormone
BiP	immunoglobulin-binding protein
BMSCs	bone marrow-derived stromal cells
BOECs	blood outgrowth endothelial cells
bp	base pairs
BPV	bovine papillomavirus
BrdU	bromodeoxy uridine
CCR5	C-C chemokine receptor type 5
CD	cytosine deaminase
cDNA	complementary DNA
CFTR	cystic fibrosis transmembrane conductance regulator
CGD	chronic granulomatous disorder
ChIP	chromatin immunoprecipitation
CIS	common integration sites
CLECs	cord-lining epithelial cells
CMV	cytomegalovirus
CoDA	context-dependent assembly
CPDG2	carboxypeptidase-G2
Cre	cyclization recombinase
CTL	cytotoxic T-lymphocyte
DAB	3,3'-diaminobenzidine
DAPI	4,6-diamino-2-phenylindole
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBD	DNA-binding domain
DMEM	Dulbecco's Modified Eagle's Medium
ds	double-stranded
DSDB	double-stranded DNA break
EBNA	EBV nuclear antigen 1

EBV	Epstein Barr Virus
EDTA	ethylenediaminetetraacetic acid
EF1α	elongation factor 1 alpha
EGFP	enhanced green fluorescent protein
EIV	equine immunodeficiency virus
EPCs	endothelial progenitor cells
EPO	erythropoietin
ER	endoplasmic reticulum
ES cells	embryonic stem cells
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FDA	Food and Drug Administration (USA)
FEIBA	Factor Eight Inhibitor Bypassing Activity
FISH	fluorescence *in situ* hybridization
FIV	feline immunodeficiency virus
FIX	coagulation factor IX
FKBP	FK-506 binding protein
FLASH	fast ligation-based automatable solid-phase high throughput
Flp	flippase
FRT	flippase recognition target
FVIII	coagulation factor VIII
GAPDH	glyceraldehydes-3-phosphate dehydrogenase
G-CSF	granulocyte colony-stimulating factor
GCV	gancyclovir
hAT	hobo/Activator/Tam3
HAT	hypoxanthine-aminopterin-thymidine
HC	high capacity
HDR	homology-directed repair
hFer	human ferritin
HFV	human foamy virus
HGPRT	hypoxanthine-guanine phosphoribosyl transferase
HIV	human immunodeficiency virus
HMWK	high molecular weight kininogen
HR	homologous recombination
HSCs	hematopoietic stem cells
HSCT	hematopoietic stem cell therapy
HSPG	heparan sulphate proteoglycans
HSV	herpes simplex virus
HSV-TK	herpes simplex virus 1 thymidine kinase
IL2RG	interleukin 2 receptor gamma
indel	insertion/deletion
iPSCs	induced pluripotent stem cells
ITR	inverted terminal repeat
kb	kilo bases
LAM-PCR	linear amplification PCR

LDL	low-density lipoprotein
LEDGF	lens epithelium-derived growth factor
LMAN1	lectin mannose-binding 1
LMO2	LIM domain only-2
LM-PCR	ligation-mediated PCR
LPL	lipoprotein lipase
LRP	lipoprotein receptor-related protein
LTR	long terminal repeat
MBE	mobile genetic elements
MEME	Multiple EM for Motif Elicitation
MLV	murine leukemia virus
MoMuLV	Moloney murine leukemia virus
mRNA	messenger RNA
MSCs	mesenchymal stem cells
NAT	nucleic acid testing
NCBI	National Center for Biotechnology Information (USA)
NHEJ	non-homologous end joining
NIH	US National Institutes of Health
NOD	non-obese diabetic
NOD-SCID	non-obese diabetic severe combined immunodeficient
NR	nitroreductase
NSG	NOD-SCID gamma
nvCJD	new variant Creutzfeldt-Jakob disease
OBA	Office of Biotechnology Activity
OH	obligate heterodimer
OPEN	oligomerized pool engineering
OTC	ornithine transcarbamylase
PAGE	polyacrylamide gel electrophoresis
PB	Piggy Bac
PBD	protein-binding domain
PBS	phosphate-buffered saline
PEG	polyethylene glycol
PNP	purine nucleoside phosphorylase
rAAV	recombinant adeno-associated virus
RBE	rep binding element
rDNA	ribosomal DNA
RFLP	restriction fragment length polymorphism
RMCE	recombinase-mediated cassette exchange
RPA	replication protein A
RPE65	retinal pigment epithelium-specific 65 kDa protein
rpm	revolutions per minute
RSV	Rous sarcoma virus
RTCGD	retroviral tagged cancer gene database
RT-PCR	reverse-transcription polymerase chain reaction
RVD	repeat-variable di-residues

S/MAR	scaffold/matrix attachment region
SAF	scaffold attachment factor
sALP	shrimp alkaline phosphatase
SB	Sleeping Beauty
SCID	severe combined immunodeficiency
SDSA	synthesis-dependent strand annealing
SIN	self-inactivating
SIV	simian immunodeficiency virus
SKY	spectral karyotyping
SP1	specificity protein 1
ss	single-stranded
SSC	saline sodium citrate
SSR	site-specific recombinase
SV40	Simian vacuolating virus 40
TAE	tris acetate EDTA
TALE	transcription activator-like effector
TALENs	transcription activator-like effector nucleases
TBE	tris borate EDTA
TEMED	tetramethylethylenediamine
TF	tissue factor
TMB	tetramethylbenzidine
TSS	transcription start sites
UCSC	University of California Santa Cruz
UTR	untranslated region
UV	ultra violet
VLDL	very-low-density lipoprotein
VWF	vonWillebrand factor
WAS	Wiskott Aldrich syndrome
YACs	yeast artificial chromosomes
ZF	zinc finger
ZFNs	zinc finger nucleases
ZFP	zinc finger protein

LIST OF PUBLICATIONS

1.   **Jaichandran Sivalingam** & Oi Lian Kon. Recent Advances and improvements in the biosafety of gene therapy. Gene Therapy – Developments and Future Perspectives. InTech, June 2011. ISBN 978-953-307-617-1.

2.   **Sivalingam J**, Krishnan S, Ng WH, Lee SS, Phan TT, Kon OL. Biosafety assessment of Site-directed transgene integration in human umbilical cord-lining cells. Mol Ther. 2010 July; 18(7): 1346-1356.

3.   Chen NK, Wong JS, Kee IH, Lai SH, Thng CH, Ng WH, Ng RT, Tan SY, Lee SY, Tan ME, **Sivalingam J**, Chow PK, Kon OL. Nonvirally modified autologous primary hepatocytes correct diabetes and prevent target organ injury in a large preclinical model. PLoS One. 2008; 3(3): e1734.

4.   **Jaichandran S**, Yap ST, Khoo AB, Ho LP, Tien SL, Kon OL. In vivo liver electroporation: Optimization and demonstration of therapeutic efficacy. Hum Gene Ther. 2006 Mar; 17(3):362-375.

5.   Chen NK, **Sivalingam J**, Tan SY, Kon OL. Plasmid-electroporated primary hepatocytes acquire quasi-physiological secretion of human insulin and restore euglycemia in diabetic mice. Gene Ther. 2005 Apr; 12(8):655-667.

6.   Long YC, **Jaichandran S**, Ho LP, Tien SL, Tan SY, Kon OL. FVIII gene delivery by muscle electroporation corrects murine hemophilia A. J Gene Med. 2005 Apr; 7(4):494-505.

Chapter 1

Introduction and literature Review

1.1.   Gene Therapy

Progress in understanding the cellular and molecular bases of human health and disease in recent decades has spawned research in the fields of regenerative medicine and gene-based therapies. These novel approaches to medical treatments offer new possibilities for mitigating, and even curing, a plethora of medical conditions ranging from rare inherited monogenic disorders, metabolic diseases, infections and even complex disorders such as cancer.

In a simplified form, gene-based therapy can be defined as any procedure aimed at genetically altering or modifying cells or tissues with exogenous genetic materials that encompasses RNA, DNA and even oligonucleotides. These molecules may be directly delivered *in vivo* into patients, often with the goal of targeting particular tissues (or organs).   Alternatively, a patient's cells may be isolated, expanded and modified *ex vivo* before reimplantation into the same subject (**Figure 1.1.1**).
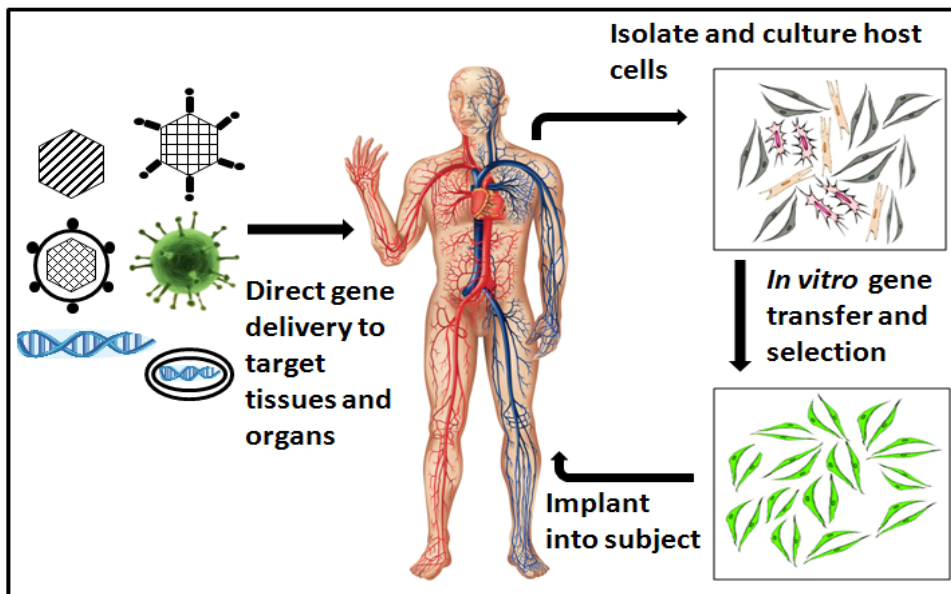
Figure 1.1.1   Gene-based therapy. Left: *In vivo* administration of vector to modify cells in target organs or tissues directly. **Right**: *Ex vivo* modification of primary somatic cells that are reimplanted into the same subject (autologous cell therapy). Cell therapy may also be allogeneic. Viral or non-viral vectors are used to deliver transgenes.

Gene-based therapy was initially conceptualized for monogenic disorders such as adenosine deaminase (ADA), alpha-1-antitrypsin, ornithine transcarbamoylase (OTC) and clotting factor (factors VIII and IX) deficiencies. These

were considered ideal candidates as reconstitution of the missing protein in each case should alleviate or abolish the disease phenotype. Conventional gene therapy has since evolved into a range of approaches that increasingly encompass cell therapy. The therapeutic spectrum of gene-and cell-based applications now extends to every area of molecular medicine to include restoration of cellular and metabolic functions in various diseases, immuno-reconstitution of tumor cells in cancer immunotherapy, targeted cancer cell ablation in suicide gene therapy, treatment of infectious diseases, genetic manipulation and reprogramming of cancer and stem cell fate, reversing degenerative vascular and brain disorders, to name just a few.

1.1.1. Historical perspectives

The original conceptualization of treating diseases by genetic engineering dates back to the 1940s when Avery, MacLeod and McCarthy pioneered the notion and demonstrated that genes could be transferred in the form of nucleic acids[1]. Gene transfer *via* viruses was first demonstrated in *Salmonella* in the early 1950s[2], and in 1962 by Szybalski in mammalian cells[3]. Early visionary investigators such as Tatum[4] envisioned "that viruses will be effectively used for man's benefit, in theoretical studies in somatic-cell genetics and possibly in genetic therapy". Indeed, virally transformed cells provided early indications of the feasibility of modifying somatic cell genomes. Replication of Rous sarcoma virus (RSV) in transformed cells and the presence of integrated simian virus 40 (SV40) viral DNA in SV40-transformed cells provided evidences for transmission of viral genes in eukaryotic cells[5].

Despite the very inefficient and crude methods of gene transfer of the 1960s, several groups were able to show that it was indeed possible to alter cellular phenotypes by transfer of whole cell genetic materials or isolated genes. For instance, Weisberger provided one of the earliest demonstrations of genetic correction of a disease phenotype when mRNA from normal bone marrow was incubated with immature erythrocytes of sickle cell subjects[6].

The early exposition of gene therapy, whilst still in its infancy, rather quickly led to premature exploits at gene therapy in human subjects that were sometimes controversial. The earliest experimentation of gene delivery in humans was carried out controversially by Rogers and collaborators who attempted to treat three patients with arginase deficiency using Shope papilloma virus[7]. A second unapproved and equally criticized human gene therapy trial was conducted in 1981 by Cline and co-workers who infused thalassemic patients with autologous bone marrow cells that had been transfected with the normal β-globin gene by the calcium phosphate method[7].

Both trials did not provide positive evidence of therapeutic efficacy although the theoretical basis of gene replacement treatment was eventually proven to be correct. What transpired after these two unapproved trials was the formation of the U.S. National Institutes of Health (NIH) Recombinant DNA Advisory Committee to regulate and approve all future human gene therapy clinical trials.

The years that followed from 1970s to 1990s saw tremendous advances in recombinant DNA technology which enabled the identification, isolation and cloning of disease related genes. Concomitant progress in viral vector development and gene delivery techniques eventually made gene transfer and expression of therapeutic genes a routine task in the laboratory. Seminal findings by the groups led by Temin[8], Scolnick[9] and Weinberg[10] resulted in the first generation viral vectors. These were retroviruses that could be augmented with therapeutically relevant genes and efficiently infected mammalian cells to induce stable expression of genes of interest. Such advances paved the way for the first approved human gene therapy clinical trial in 1990 for children with ADA deficiency[11]. By the turn of the millennium, almost 4000 patients had been enrolled in more than 500 gene therapy clinical trials worldwide[5], alas with variable and generally limited success. Nonetheless, these studies have contributed to the field by highlighting several areas for improvement and emphasizing the need to refine treatment methods to mitigate risks to patients. Initial high hopes that gene therapy could be readily translated into the standard clinical practice has yet to be fulfilled, partly because enthusiasm from a few clinical successes has been marred by the occurrence of adverse and serious iatrogenic complications in a small number of trial subjects[12]. Experience of these unexpected complications has reiterated the need to understand and evaluate genotoxicity and other risks of any given gene therapy approach, and for pertinent biosafety improvements to be incorporated into proposed treatment modalities.

1.1.2. Approaches for gene therapy

Gene delivery can be achieved using either viral vectors or non-viral vectors. The latter may be episomally maintained or integrated into the host genome. To date, five main classes of viral vectors have been tested for clinical applications. These include retroviruses, adenoviruses, adeno-associated viruses (AAV), lentiviruses and herpes simplex viruses (HSV)[13]. Non-viral vectors most often utilize plasmid DNA which can be delivered into cells or tissues by physical methods such as electroporation, gene-gun bombardment, sonoporation, hydrodynamic injection or by chemical methods that utilize calcium phosphate, polymeric carriers, cell-penetrating

peptides, cationic and anionic lipids[14]. Generally, gene therapy can be segregated into two main categories depending on whether the therapeutic gene is directly delivered *in vivo* to transduce or transfect cells in target organs and tissues; or whether specific cell types (allogeneic or autologous) are first expanded and gene-modified *ex vivo* before implantation into patients.

1.1.2.1. *In vivo* gene delivery

Transgene delivery by viral vectors is the most widely attempted gene therapy approach and exploits the natural process of infection by viral pathogens. Certain viral infections result in delivery and expression of viral genes in infected host cells. Pioneering scientists recognized this trait and envisioned that genetically modified viruses could be utilized to deliver and express therapeutic transgenes. This approach has several appealing characteristics namely, its non-invasive nature, the simplicity of direct vector delivery and known tissue-specific transduction efficiency of some viral vectors. Thus, it is no surprise that a majority of current *in vivo* gene delivery techniques involve intravascular delivery of viral vectors to achieve tissue-specific (e.g. liver, muscle) or generalized effects. Successful demonstration of *in vivo* viral vector delivery for correction of genetic disorders thus far include adenoviral delivery for correction of OTC deficiency[15], intramuscular delivery of rAAV for clotting factor IX (FIX) deficiency[16] and intramuscular delivery of rAAV1 for lipoprotein lipase deficiency[17].

In vivo gene delivery has also been attempted with non-viral vectors employing hydrodynamic delivery of naked or chemically modified plasmid DNA, localized or systemic delivery of plasmid DNA in combination with physical gene transfer methods such as electroporation or sonoporation, aerosol-mediated plasmid delivery to the lungs and nanoparticle-mediated plasmid delivery to target tissues, to name just a few. While *in vivo* delivery of non-viral vectors has not been employed in clinical trials as extensively as viral vectors owing to low efficiencies of gene transfer, a notable example is localized transdermal injection of plasmid DNA combined with electroporation for eradication of solid tumours[18].

The main issues with regard to *in vivo* gene delivery can be summarized as follows. Firstly, is the issue of immunogenicity of some *in vivo* administered vectors. Secondly, *in vivo* gene delivery may result in the transduction of multiple cell types, including cells of the immune system, especially when delivered systemically. Unintended transduction of cells may be undesirable for certain gene therapy applications, such as tumor-targeted suicide gene delivery where stringent targeting

of therapeutic agents to cancerous cells is crucial. Where the inadvertent transduction is to cells of the immune system, there could be potential heightened risks of immune response towards the transgene product[19]. Thirdly, persistent concerns of germ-line transduction remain despite reports that have dismissed previously reported evidence for germ-line transduction in animals and even human subjects[20]. Lastly, the known risks for insertional mutagenesis and oncogenesis is a major reservation with the use of integrating vectors. Another caveat to *in vivo* delivery of vectors is that very little can be done to interrupt or abort gene delivery once the vector has been administered *in vivo*. Taken together, these concerns mandate long-term biosafety studies in animal models to evaluate oncogenicity and other adverse complications before embarking on clinical trials of *in vivo* gene delivery.

1.1.2.2. *Ex vivo* cellular therapy

The *ex vivo* cell therapy approach relies on derivation, *ex vivo* culture, gene modification and re-administration of patient-derived somatic cells to serve either as carriers for expressing a deficient protein, or for repairing or regenerating damaged tissues or organs (regenerative medicine). Successful engraftment followed by long-term survival of implanted cells and durable transgene expression could, in theory, achieve long-term correction of the disease phenotype. Not restricted to the foregoing description, *ex vivo* cell therapy also encompasses cancer immunotherapy whereby genetically modified patient- or donor-derived cells are used as immunogens to educate and activate the immune system to eliminate cancerous cells. In suicide gene therapy, normal cells are modified to home in and kill cells within the tumor by expressing a gene product that converts a systemically administered prodrug into a cytotoxic agent.

Over the past two decades, many *ex vivo* cell therapy approaches have been developed to correct genetic or metabolic deficiencies, the best studied and most successful of which are based on hematopoietic stem cells (HSCs) transplantations. Successful correction of various life-threatening diseases of the blood and immune systems such as ADA-severe combined immunodeficiency (SCID)[21], *IL2RG*-SCID[22] (also known as SCID-X1), chronic granulomatous disorder[23] (CGD), Wiskott-Aldrich syndrome[24](WAS) have been reported with transplantation of genetically modified HSCs. Beside HSCs, other somatic and stem cell types are also under active investigation as cellular carriers for correcting monogenic disorders. Primary human fibroblasts[25] and endothelial progenitor cells[26] (EPCs) have been used to correct clotting factor deficiencies (FVIII, FIX), epidermal and limbal stem cells have been

used for ocular gene therapy[27] and mesenchymal stem cells (MSCs)[28] have found applications in regenerative therapy aimed at repairing defects of the heart, bones, ligaments and joints. More recent developments point to the potential of embryonic stem cells and induced pluripotent stem cells (iPSCs) for both cell-based gene therapies and regenerative medicine.

The *ex vivo* cell therapy approach has several advantages over *in vivo* vector delivery. Subjects are not directly exposed to gene therapy vectors with the attendant risks of adverse immune responses and germ-line transmission. Targeted modification of specific cell types is much more feasible as is the potential to comprehensively evaluate the biosafety of genome-modified cells before *in vivo* implantation. Important factors to consider when developing *ex vivo* cell-based therapy include the ease of isolating and expanding primary cells *ex vivo* to derive clinically meaningful cell numbers. The choice of cell type would depend upon the disease to be treated, and could often be the same or similar cell types that are affected by the disease. The availability and amenability of a particular cell type for *ex vivo* derivation, culture and gene modification as well as the ability to efficiently express or secrete a properly processed protein are key considerations in the choice of cell type. The choice of vectors for *ex vivo* cell modification is another important consideration. Vectors that integrate the transgene into the genome could theoretically allow for durable transgene expression that is desirable for certain diseases and essential for cells (e.g. HSCs) that replicate *in vivo*. Although most current approaches employ integrating viral vectors i.e. retro- and lentiviral vectors, non-viral integrating systems such as site-specific recombinases (transposons and phiC31 integrase) and site-specific nucleases (ZFNs, TALENs and meganucleases) have come to the fore recently as safer strategies[29]. The known occurrence of insertional mutagenesis and oncogenic transformation is arguably the greatest challenge in clinical adoption of most integrating systems, whether viral or non-viral[30]. Non-integrating or episomally maintained vectors may be safer and could also allow for durable transgene expression in post-mitotic cells or even in dividing cells provided the vector is episomally maintained within replicating cells. Non-integrating vectors have yet to be successfully adapted for *ex vivo* cell therapy. Non-integrating viral vectors such as adenoviral and rAAV vectors have typically been used for *in vivo* gene delivery while non-viral episomal vectors have only been tested *in vitro* or in small animal models.

1.1.3. Vectors used in gene therapy

1.1.3.1. Viral vectors

Viral vectors are mainly used in current research to develop clinical gene therapy. The generally higher efficiencies of gene transfer *in vitro* and *in vivo*, and the wide range of cell types that can be efficiently transduced make viral vectors the agents of choice for most laboratory and clinical studies. The major classes of viruses that have found application in gene therapy include the gamma retrovirus, HIV-lentivirus, HSV, adenovirus and rAAV[13]. These viruses can be further sub-categorized as non-integrating (adenovirus, rAAV, HSV, integrase-deficient lenti- and retroviruses) and integrating (retrovirus, lentivirus) vectors[31].

1.1.3.1.1.   Non-integrating viral vectors

Non-integrating viral vectors are predominantly maintained within host cell nuclei as episomes and rarely integrate into the host genome. They are especially useful for transfecting post-mitotic cells where vector loss as a consequence of cell division is less of a concern. Adenovirus, rAAV and HSV are commonly used non-integrating viral vectors.

Adenoviral vectors are a class of non-enveloped viral vectors with broad tissue tropism and are capable of transducing non-dividing as well as proliferating cells such as macrophages, fibroblasts, muscle, liver, neural cells and tumor cell lines[31]. They are predominantly maintained as episomes within the host cell nucleus and as such are considered safer than integrating vectors. Adenoviral vectors have been demonstrated to mediate persistent transgene expression in non-dividing cells. Gutless or helper dependent adenoviral vectors are improved versions of the first generation vectors in which most viral genes have been deleted, rendering them replication-incompetent, less immunogenic and toxic, and capable of accommodating DNA inserts of up to 37 kb. Adenoviral vectors have been used extensively for brain, muscle, lung and liver gene therapy studies with successful outcomes in several pre-clinical studies for genetic disorders such as for cystic fibrosis[32], hemophilia B[33] and Crigler-Najjar syndrome[34]. They have applied for anti-cancer effects such as tumor ablation by adenoviral-mediated expression of suicide genes[35] and expression of recombinant tumor suppressor, p53[36]. A major drawback of adenoviral vectors *in vivo* is the stimulation of innate immune responses directed at viral capsid proteins. Several strategies to counter this adverse complication, such as transient immunosuppression, modification of capsid proteins and viral dosage optimization, are being tested. .

Recombinant AAV vectors[37] are another widely utilized class of vectors that have been shown to successfully transduce a variety of tissues such as muscle, brain,

lung, liver and retina. These single stranded DNA viruses, belonging to the genus *Dependovirus*es, require expression of viral replication and assembly proteins from helper viruses for assembly and have a packaging capacity of around 5 kb. AAV vectors are capable of transducing dividing and non-dividing cells. However, unlike wild type AAV vectors which have a predilection for integrating into human chromosome 19, rAAV are maintained predominantly as episomes, although random integration has also been reported, albeit at very low frequencies[38]. Durable transgene expression following rAAV transduction has been found in post-mitotic tissues, such as muscle and liver[39]. The capsid protein sequences of the viruses define their serotype and determine different tropisms for specific cell types. Although approximately 110 primate AAV capsid sequences are known, thus far there are only 12 commonly reported serotypes, AAV1 to AAV12[37]. AAV2, by far the most frequently utilized serotype, has been used successfully to transduce skeletal muscle, neurons and liver cells and was the chosen vector in several clinical trials, such as for expression of *RPE65* gene in retina as treatment for Leber's congenital amaurosis[27], FIX expression in liver for hemophilia B[16] and cystic fibrosis transmembrane conductance regulator (*CFTR*) expression in lung for cystic fibrosis[40], to name just a few. Other AAV serotypes have also met with success in transducing specific cell types e.g. AAV6 for airway epithelial cells[41], AAV1 and AAV5 for vascular endothelial cells[42] and AAV8 for hepatocytes[43]. In clinical trials, some efficacy was achieved by AAV1-mediated transduction of muscle to correct α1-antitrypsin[44] and lipoprotein lipase (*LPL*) deficiencies[17], while encouraging results using AAV8 for liver transduction in hemophilia B patients was reported recently[16]. At present, humoral immunity against viral capsid proteins which has often reduced the effectiveness of rAAV-mediated gene transfer *in vivo* remains a major challenge to be resolved.

Herpes simplex virus-1 (HSV-1)-derived vectors are another class of non-integrating vectors[45]. Though less commonly used, they have specialized utility in oncolytic cancer gene therapy, neurological research and treatment of neurological disorders, given their natural neurotropism. HSV-1 vectors have the advantage of a large cloning capacity (40 – 150 kb) and may be engineered for latency in infected cells to achieve durable transgene expression. Conversely, wild type HSV-1 with a lytic life cycle could be beneficial for infecting and killing tumor cells. Current limitations of the HSV vectors are low *in vivo* transduction efficiencies, transient transgene expression and cellular toxicity.

Integrase-deficient retroviral[46] and lentiviral[47] vectors can be considered as yet another class of non-integrating viral vectors. Engineered mutations of the viral integrase protein has yielded vectors with a highly reduced ability to integrate into the host genome. While a high percentage of such vectors is episomal in transduced cells, residual integrase activity may still enable vector integration into the genome.

1.1.3.1.2.   Integrating viral vectors

To date, the most promising results in clinical trials have been attained using integrating viral vectors. Their propensity to stably integrate transgenes into the genome ensures stable maintenance of the transgene in actively dividing cells and potentially favors durable transgene expression thereby circumventing the need for repeated vector administration. Retroviral and lentiviral vectors are the major types of integrating vectors that have produced evidence in efficacy in pre-clinical studies and clinical trials. However, a major caveat is the risk of insertional mutagenesis given the propensity of certain integrating viral vectors to integrate preferentially into actively transcribed genes and close to transcription start sites[30].

Retroviruses are one of the earliest developed and extensively utilized integrating gene transfer vectors[48]. They are enveloped single-stranded RNA viruses that are reverse transcribed into double-stranded DNA in infected cells. Retroviruses integrate randomly in the genome, with a predilection for transcription start sites. Deletion of the viral *gag*, *pol* and *env* genes has enabled carriage of transgenes up to 7.5 kb. The gag, pol and env proteins essential for viral assembly must be provided in *trans* by packaging and helper cell lines. Given that the nuclear membrane must break down for retrovirus entry, only dividing cells can be transduced by retroviral vectors. For gene therapy applications, replication-defective retroviruses are used and have been shown to effectively transduce a wide variety of actively dividing cells, such as fibroblasts, hepatocytes, bone marrow MSCs and HSCs. Most retroviral vectors are derivatives of the Moloney murine leukemia virus (MoMuLV), while others are modifications of RSV, avian sarcoma-leukosis virus (ASLV), human immunodeficiency virus (HIV) and human foamy virus (HFV). Retroviral vectors have been used with success in gene therapy clinical trials for several monogenic disorders e.g. SCID-ADA, SCID-*IL2RG*, WAS and CGD, although with regrettable iatrogenic adverse events in a few patients in trials for SCID-*IL2RG*[49, 50], CGD[51] and WAS (http://www.asgct.org/media/news-releases/?c=505). As with most integrating vectors, the major disadvantage of retroviral vectors is the risk of inducing insertional mutagenesis due to their aforementioned pattern of integration. More recent studies of

foamy retroviruses appear to suggest a safer integration profile, *viz.* reduced frequency of integration near transcription start sites[52]. Self-inactivating (SIN)-retroviral vectors are being investigated as a means of reducing risks of insertional activation of neighboring genes[53].

Lentiviruses have an RNA genome and are a retrovirus subclass. They efficiently infect a wide variety of cell types and have an integration profile quite dissimilar to retroviruses in that they do not have a predilection for transcription start sites, although they preferentially integrate into active transcription units[54]. Unlike retroviral vectors, lentiviral vectors will transduce both dividing and non-dividing cells to deliver transgenes of up to 8 kb. Most lentiviral vectors are developed from human immunodeficiency virus type -1 (HIV-1) although others are derivatives of HIV-2, simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV) and equine immunodeficiency virus (EIV)[55]. Lentiviral vectors have been used in several gene therapy clinical trials, thus far without reports of serious adverse events[56]. Successful clinical outcomes have been reported in lentiviral-mediated gene therapy of HIV[57], X-linked adrenoleukodystrophy[58], β-thalassemia[59], among others. The development of SIN-lentiviral vectors have strengthened the safety profiles of these vectors and have increasingly advocated their replacement of retroviral vectors in gene therapy studies[60]. New clinical studies for SCID-X1, WAS and SCID-ADA using SIN-lentiviral vectors instead of retroviral vectors are currently ongoing or being initiated (clinical trials.gov). Despite the anticipated superior biosafety of lentiviral vectors, it should be noted that this is currently speculative because risks of insertional mutagenesis and inactivation of crucial genes or tumor suppressors remain, owing to their propensity to integrate randomly in active transcription units.

1.1.3.2. Non-viral vectors

Non-viral vectors are usually naked or modified plasmid DNA that are delivered to cells by chemical or physical transfection methods. Chemical transfection methods include the use of calcium phosphate, polyethylene glycol (PEG), cationic or anionic lipid reagents that conjugate to DNA and bring about their internalization, transit to the nucleus culminating in transgene expression by transfected cells. Physical methods of plasmid delivery include gene-gun bombardment, sonoporation, electroporation, and hydrodynamic gene delivery that deliver DNA into cells and eventually to the nucleus by invoking incompletely understood physical changes to the cell and/or nuclear membranes.

Plasmid DNA thus delivered is predominantly episomal in the transfected cell nucleus, unless co-delivered with a protein (such as a recombinase) that induces genomic integration of the plasmid. On the other hand, non-integrating plasmid DNA does not replicate in tandem with mitosis, with the exception of plasmids modified for episomal maintenance by tethering to host chromosomes and replicating with successive cell divisions[61]. As the work reported in this thesis used only non-viral vectors, they will be discussed in greater depth in the following sections.

1.1.3.2.1.   Episomal non-viral vectors

One of the most apparent advantages of extra-chromosomal vectors as gene transfer agents is the exponentially decreased risks of insertional mutagenesis compared to integrating vectors. Episomal plasmids can be maintained at high copy number, have potentially higher levels of transgene expression and are less likely to undergo transgene silencing or exhibit positional variegation effects associated with genomic integrations[61].

The essential characteristics of extra-chromosomal vectors are episomal maintenance, autonomous replication and segregation into daughter cells. Episomal vectors can be categorized as either virus-based if they rely on viral origins of replication and other virally encoded proteins for replication and partitioning into daughter cells, or chromosome-based, if they depend on genomic elements. Examples of virus-based episomal vectors include those that use viral replicons of SV40, bovine papillomavirus (BPV) and Epstein-Barr virus (EBV) or that carry limited viral components such as oriP/EBV nuclear antigen 1 (EBNA1). Chromosome-based episomal vectors include the scaffold/matrix attachment region (S/MAR)-based pEPI vectors and artificial chromosomes[61].

Episomal non-viral vectors represent a class of vectors that could serve as efficient and safe gene therapy agents for long-term expression not only in *ex vivo* modified cells but also *in vivo* in tissues. The exciting possibility of utilizing them for gene transfer into adult and embryonic stem cells *ex vivo* as well as their superior biosafety warrants their investigation for clinical applications. However, caution must be exercised and, as with all other modalities, the minute but potential risk of random vector insertion[62] must be evaluated and its likely implications studied.

1.1.3.2.2.   Integrating non-viral vectors

Unlike integrating viral vectors, most non-viral vectors have no intrinsic capacity for genome integration except for rare random events, usually in genomic

break sites. However, genomic integration of non-viral vectors may be greatly enhanced by co-expression of bacteriophage- or virus-derived recombinases or integrases. Several well-studied recombinases and integrases are known to recognize distinct DNA motifs in both the genome as well as in non-viral vectors, and thereby mediate site-specific integration of vectors into genomic target sites, although recognition specificity is often not highly stringent. Common examples of recombinases are the bacteriophage P1-derived *cre* recombinase which mediates integration of loxP-containing vectors into pre-integrated loxP sites or pseudo loxP sites within the genome, *Saccharomyces cerevisiae*-derived flippase (Flp) recombinase which similarly recognizes and integrates into flippase recognition target (FRT) sites, and the integrase derived from the *Streptomycete* bacteriophage phiC31 which integrates *attB* containing vectors into genomic *attP* or pseudo *attP* sites. There are also several transposons such as Sleeping Beauty, PiggyBac and *Tol2* that mediate transposition of inverted terminal repeats (ITRs)-flanked transgene vectors into genomic regions often defined by TA and TTAA target sequences[63, 64]. Targeted integration of transgenes specifically into the AAVS1 site of the human genome has been demonstrated with the expression of AAV Rep78 or 68 proteins and AAV ITR-flanked transgene vectors[65, 66]. Non-viral integrating systems that have generated considerable recent interest involve the use of site-specific nucleases that cleave the genome at a defined or unique region. Site-specific genomic cleavage by homing endonucleases (meganucleases), zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) enhance homologous recombination-mediated integration of exogenous DNA sequences into pre-defined genomic sites. TALENs and ZFNs are currently being investigated for their accuracy and efficiency of transgene integrations into unique and safe sites within the genome.

At present, a majority of clinical gene therapy studies are conducted with viral vectors, given their generally superior efficiency both *in vivo* and *ex vivo*. However, viral vectors are still plagued by concerns of immunogenicity, cytotoxicity and genotoxicity. Demonstrating efficient, effective and safe integration of transgenes with non-viral vectors could positively contribute to the eventual development of clinical gene therapy.

1.1.4. Clinical trials – successes and failures

As of 2013, the Journal of Gene Medicine clinical trials database recorded a total of 1843 gene therapy clinical trials, 64.4% of which were directed at cancer and related diseases (last accessed on 3rd January 2013) (**Figure 1.1.2**) . Given the greater depth of understanding of molecular virology, the broad tropism of viral vectors and

their superior efficiencies of gene transfer, transgene delivery via viral vectors has been the most favored option (66.8 % of all trials). The demographics of clinical gene therapy trials worldwide are shown in **Figure 1.1.2**.
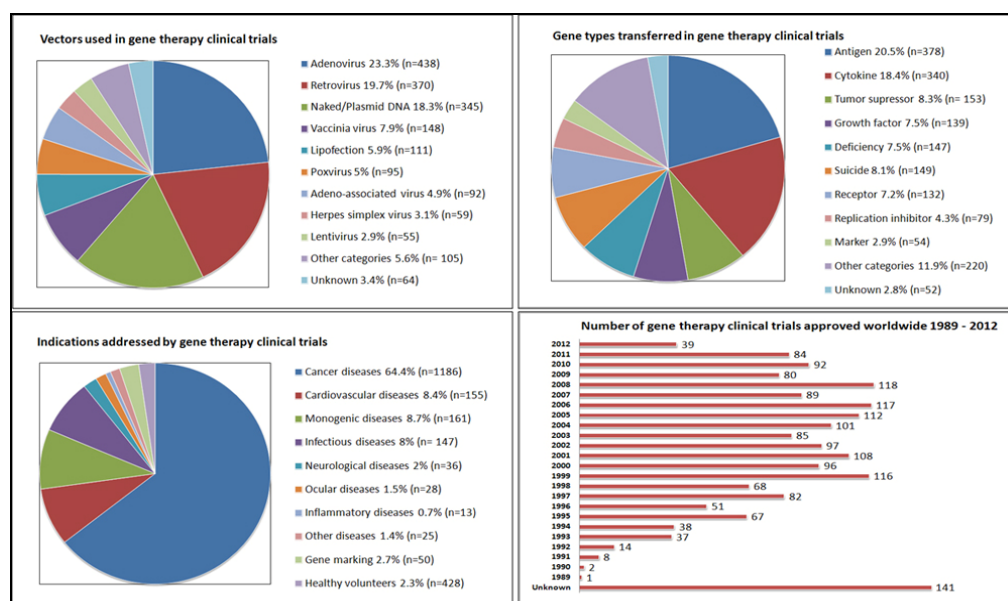
Figure 1.1.2 Characteristics of clinical gene therapy trials. Categorization of gene therapy trials according to indications, vectors used, gene types transferred and annual number of approved trials. (*Images reproduced from Journal of Gene Medicine clinical trials database (http://www.wiley.com/legacy/wileychi/genmed/clinical/)).*

Despite the impressive numbers of gene therapy trials, it is worth noting that only a small number had clinically meaningful outcomes. The first clinical success was treatment of X-linked severe combined immunodeficiency (SCID-X1)[22], a disease characterized by arrested development of the immune system due to mutations in the interleukin-2 receptor common gamma chain gene (*IL2Rγ*). Nine of ten treated patients achieved long-term immune reconstitution and marked clinical improvement following implantation of gene-modified HSCs[67]. More success stories echoed from similar clinical trials in London, U.K., of the same disorder[68]. In the years that followed, long-term therapeutic efficacy was also reported in clinical trials for another form of SCID, SCID-ADA[21, 69]. In 2006, gene therapy scored more successes when impressive results were reported in two patients treated for CGD[23], caused by inactivating mutations of gp91phox (*CYBB*) gene and characterized by neutrophil dysfunction and severe recurrent infections. More recent and notable cases of clinical success are treatment of hemophilia B[16], WAS[24], X-linked adrenoleukodystrophy (ALD)[58], Leber's congenital amaurosis[70] and Parkinson's disease (to restore dopamine expression in the subthalamic nuclei)[71].

Although these impressive clinical outcomes provided incontrovertible proof-of-principle, it soon became evident that treatment benefits could occur in tandem with significant adverse effects when serious iatrogenic complications were reported in a small number of patients. The first gene therapy death was reported in 1999 from an OTC clinical trial conducted at the University of Pennsylvania. This was ascribed to a massive immune response to the adenoviral vector used in that trial[72]. Gene therapy suffered the heaviest blows in the years 2003 to 2006, and attracted close scrutiny by regulatory authorities and the medical fraternity when five successfully treated SCID-X1 patients (from two different clinical trials) developed T-cell lymphoblastic leukemia, three to six years after treatment with autologous bone marrow-derived CD34+ hematopoietic cells transduced with a murine leukemia virus (MLV) gammaretroviral vector to express the *IL2Rγ* gene[49, 50]. Random integration of the MLV gamma retroviral vector that had strong enhancer elements in the long terminal repeat (LTR) regions resulted in insertional activation of the LIM domain only-2 (*LMO2*) proto-oncogene. This mutagenic event likely promoted clonal proliferation of T cells that culminated in acute lymphoblastic leukemia. In a different trial in 2007, Targeted Genetics Corporation was forced to halt its gene therapy trial for rheumatoid arthritis involving intra-articular injection of an adenoviral vector expressing *tgAAC94* following the death of a patient. In this case however, investigations by the US Food and Drug Administration (FDA) exonerated gene therapy as the direct cause of death[73], although there was evidence of vector-induced immune response; and the trials have since recommenced. The inherent risks of insertional mutagenesis by viral vectors surfaced again in another clinical trial in 2006 for treatment of CGD. Two adult CGD patients infused with granulocyte colony-stimulating factor (G-CSF)–mobilized peripheral blood CD34+ cells transduced with MLV gammaretroviral vector expressing *gp91phox* had markedly improved neutrophil functions and resistance to life-threatening infections. Regrettably, both subjects later developed myelodysplasia and one subject died from this complication[51]. Myelodysplasia probably developed from random integration of the gammaretroviral vector that activated the expression of a proto-oncogene, *MDS-EVI1*[51]. As of this writing, the most recent cases of adverse gene therapy outcome, brought to light by the American Society of Gene and Cell therapy (http://www.asgct.org/media/news-releases/?c=505 and ASGCT meeting May 2012) affected four of ten WAS patients treated at the Hannover Medical School using a gammaretroviral vector similar to that used in the SCID-X1 trials. These patients

were reported to have developed leukemia. Comprehensive clinical evaluations of these adverse events have yet to be disclosed.

In summary, there is clear evidence that gene therapy can be clinically effective. Moreover, it offers the only treatment for certain serious life-threatening diseases that are currently untreatable or poorly treated. An important issue that must be addressed if gene therapy is to mature from experimental treatment to clinical standard of care is that of biosafety. The occurrence of serious iatrogenic outcomes, albeit uncommon, has brought into sharp focus the inherent risks of genetic modifications.

1.1.5.　References

1. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. J Exp Med. (1944) 79:137-158.

2. Zinder N, Lederberg J. Genetic exchange in Salmonella. J Bacteriol. (1952) 64:679-699.

3. Szybalska EH, Szybalski W. Genetics of human cell line. IV. DNA-mediated heritable transformation of a biochemical trait. Proc Natl Acad Sci U S A. (1962) 48:2026-2034.

4. Tatum EL. Molecular biology, nucleic acids, and the future of medicine. Perspect Biol Med. (1966) 10:19-32.

5. Scollay R. Gene therapy: a brief overview of the past, present, and future. Ann N Y Acad Sci. (2001) 953:26-30.

6. Weisberger AS. Induction of alered globin synthesis in human immature erythrocytes incubated with ribonucleoprotein. Proc Natl Acad Sci U S A. (1962) 48:68-80.

7. Wolff J, Lederberg J. An early history of gene transfer and therapy. Hum Gene Ther. (1994) 5:469-480.

8. Shimotohno K, Temin HM. Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. Cell. (1981) 26:67-77.

9. Wei CM, Gibson M, Spear PG, Scolnick EM*, et al*. Construction and isolation of a transmissible retrovirus containing the src gene of Harvey murine sarcoma virus and the thymidine kinase gene of herpes simplex virus type 1. J Virol. (1981) 39:935-944.

10. Tabin CJ, Hoffmann JW, Goff SP, Weinberg RA, *et al*. Adaptation of a retrovirus as a eucaryotic vector transmitting the herpes simplex virus thymidine kinase gene. Mol Cell Biol. (1982) 2:426-436.

11. Blaese RM, Culver KW, Miller AD, Carter CS, *et al*. T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. Science. (1995) 270:475-480.

12. Kustikova O, Brugman M, Baum C. The genomic risk of somatic gene therapy. Semin Cancer Biol. (2010) 20:269-278.

13. Walther W, Stein U. Viral vectors for gene transfer: a review of their use in the treatment of human diseases. Drugs. (2000) 60:249-271.

14. Niidome T, Huang L. Gene therapy progress and prospects: nonviral vectors. Gene Ther. (2002) 9:1647-1652.

15. Raper SE, Wilson JM, Yudkoff M, Robinson MB, *et al*. Developing adenoviral-mediated in vivo gene therapy for ornithine transcarbamylase deficiency. J Inherit Metab Dis. (1998) 21 Suppl 1:119-137.

16. Nathwani AC, Tuddenham EGD, Rangarajan S, Rosales C, *et al*. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med. (2011) 365:2357-2365.

17. Gaudet D, Méthot J, Déry S, Brisson D, *et al*. Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPL(S447X)) gene therapy for lipoprotein lipase deficiency: an open-label trial. Gene Ther. (2012)

18. Heller LC, Heller R. Electroporation gene therapy preclinical and clinical trials for melanoma. Curr Gene Ther. (2010) 10:312-317.

19. Nayak S, Herzog RW. Progress and prospects: immune responses to viral vectors. Gene Ther. (2010) 17:295-304.

20. Favaro P, Downey HD, Zhou JS, Wright JF*, et al*. Host and vector-dependent effects on the risk of germline transmission of AAV vectors. Mol Ther. (2009) 17:1022-1030.

21. Aiuti A, Slavin S, Aker M, Ficara F*, et al*. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science. (2002) 296:2410-2413.

22. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F*, et al*. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science. (2000) 288:669-672.

23. Ott MG, Schmidt M, Schwarzwaelder K, Stein S*, et al*. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med. (2006) 12:401-409.

24. Boztug K, Schmidt M, Schwarzer A, Banerjee PP*, et al*. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N Engl J Med. (2010) 363:1918-1927.

25. Roth DA, Tawa NE, O'Brien JM, Treco DA*, et al*. Nonviral transfer of the gene encoding coagulation factor VIII in patients with severe hemophilia A. N Engl J Med. (2001) 344:1735-1742.

26. Matsui H, Shibata M, Brown B, Labelle A*, et al*. Ex vivo gene therapy for hemophilia A that enhances safe delivery and sustained in vivo factor VIII expression from lentivirally engineered endothelial progenitors. Stem Cells. (2007) 25:2660-2669.

27. Simonelli F, Maguire AM, Testa F, Pierce EA*, et al*. Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. Mol Ther. (2010) 18:643-650.

28. Wang J, Liao L, Tan J. Mesenchymal-stem-cell-based experimental and clinical trials: current status and open questions. Expert Opin Biol Ther. (2011) 11:893-909.

29. Galetto R, Duchateau P, Pâques F. Targeted approaches for gene therapy and the emergence of engineered meganucleases. Expert Opin Biol Ther. (2009) 9:1289-1303.

30. Baum C. Current opinions in hematology 2007 vol 14 Insertional mutagenesis in gene therapy. Hematology. (2007) 14:337-342.

31. Brunetti-Pierri N, Ng P. Progress and prospects: gene therapy for genetic diseases with helper-dependent adenoviral vectors. Gene Ther. (2008) 15:553-560.

32. Griesenbach U, Geddes DM, Alton EWFW. Gene therapy progress and prospects: cystic fibrosis. Gene Ther. (2006) 13:1061-1067.

33. Brunetti-Pierri N, Liou A, Patel P, Palmer D, *et al*. Balloon Catheter Delivery of Helper-dependent Adenoviral Vector Results in Sustained, Therapeutic hFIX Expression in Rhesus Macaques. Mol Ther. (2012) 20:1863-1870.

34. Toietta G, Mane VP, Norona WS, Finegold MJ, *et al*. Lifelong elimination of hyperbilirubinemia in the Gunn rat with a single injection of helper-dependent adenoviral vector. Proc Natl Acad Sci U S A. (2005) 102:3930-3935.

35. Sharma A, Tandon M, Bangari DS, Mittal SK, *et al*. Adenoviral vector-based strategies for cancer therapy. Current drug therapy. (2009) 4:117-138.

36. Swisher SG, Roth JA, Komaki R, Gu J, *et al*. Induction of p53-regulated genes and tumor regression in lung cancer patients after intratumoral delivery of adenoviral p53 (INGN 201) and radiation therapy. Clin Cancer Res. (2003) 9:93-101.

37. Schultz BR, Chamberlain JS. Recombinant adeno-associated virus transduction and integration. Mol Ther. (2008) 16:1189-1199.

38. Nowrouzi A, Penaud-Budloo M, Kaeppel C, Appelt U, *et al*. Integration frequency and intermolecular recombination of rAAV vectors in non-human primate skeletal muscle and liver. Mol Ther. (2012) 20:1177-1186.

39. Alexander IE, Cunningham SC, Logan GJ, Christodoulou J, *et al*. Potential of AAV vectors in the treatment of metabolic disease. Gene Ther. (2008) 15:831-839.

40. Moss RB, Rodman D, Spencer LT, Aitken ML*, et al*. Repeated adeno-associated virus serotype 2 aerosol-mediated cystic fibrosis transmembrane regulator gene transfer to the lungs of patients with cystic fibrosis: a multicenter, double-blind, placebo-controlled trial. Chest. (2004) 125:509-521.

41. Halbert CL, Allen JM, Miller AD. Adeno-associated virus type 6 (AAV6) vectors mediate efficient transduction of airway epithelial cells in mouse lungs compared to that of AAV2 vectors. J Virol. (2001) 75:6615-6624.

42. Chen S, Kapturczak M, Loiler SA, Zolotukhin S*, et al*. Efficient transduction of vascular endothelial cells with recombinant adeno-associated virus serotype 1 and 5 vectors. Hum Gene Ther. (2005) 16:235-247.

43. Wang L, Bell P, Lin J, Calcedo R*, et al*. AAV8-mediated hepatic gene transfer in infant rhesus monkeys (Macaca mulatta). Mol Ther. (2011) 19:2012-2020.

44. Brantly ML, Chulay JD, Wang L, Mueller C*, et al*. Sustained transgene expression despite T lymphocyte responses in a clinical trial of rAAV1-AAT gene therapy. Proc Natl Acad Sci U S A. (2009) 106:16363-16368.

45. Marconi P, Argnani R, Epstein AL, Manservigi R*, et al*. HSV as a vector in vaccine development and gene therapy. Adv Exp Med Biol. (2009) 655:118-144.

46. Yu SS, Dan K, Chono H, Chatani E*, et al*. Transient gene expression mediated by integrase-defective retroviral vectors. Biochem Biophys Res Commun. (2008) 368:942-947.

47. Banasik MB, McCray PB. Integrase-defective lentiviral vectors: progress and applications. Gene Ther. (2010) 17:150-157.

48. VandenDriessche T, Collen D, Chuah MKL. Biosafety of onco-retroviral vectors. Curr Gene Ther. (2003) 3:501-515.

49. Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C*, et al*. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest. (2008) 118:3143-3150.

50. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, *et al*. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest. (2008) 118:3132-3142.

51. Stein S, Ott MG, Schultze-Strasser S, Jauch A, *et al*. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med. (2010) 16:198-204.

52. Erlwein O, McClure MO. Progress and prospects: foamy virus vectors enter a new age. Gene Ther. (2010) 17:1423-1429.

53. Yi Y, Hahm SH, Lee KH. Retroviral gene therapy: safety issues and possible solutions. Curr Gene Ther. (2005) 5:25-35.

54. Mitchell RS, Beitzel BF, Schroder ARW, Shinn P, *et al*. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. (2004) 2:E234.

55. Debyser Z. Biosafety of lentiviral vectors. Curr Gene Ther. (2003) 3:517-525.

56. Schambach A, Baum C. Clinical application of lentiviral vectors - concepts and practice. Curr Gene Ther. (2008) 8:474-482.

57. Manilla P, Rebello T, Afable C, Lu X, *et al*. Regulatory considerations for novel gene therapy products: a review of the process leading to the first clinical lentiviral vector. Hum Gene Ther. (2005) 16:17-25.

58. Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, *et al*. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science. (2009) 326:818-823.

59. Kaiser J. Gene therapy. Beta-thalassemia treatment succeeds, with a caveat. Science. (2009) 326:1468-1469.

60. Modlich U, Navarro S, Zychlinski D, Maetzig T*, et al*. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther. (2009) 17:1919-1928.

61. Lufino MMP, Edser PAH, Wade-Martins R. Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. Mol Ther. (2008) 16:1525-1538.

62. Wang Z, Troilo PJ, Wang X, Griffiths TG*, et al*. Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. Gene Ther. (2004) 11:711-721.

63. Wilson MH, Coates CJ, George AL. PiggyBac transposon-mediated gene transfer in human cells. Mol Ther. (2007) 15:139-145.

64. Vigdal TJ, Kaufman CD, Izsvák Z, Voytas DF*, et al*. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. J Mol Biol. (2002) 323:441-452.

65. Lamartina S, Roscilli G, Rinaudo D, Delmastro P*, et al*. Lipofection of purified adeno-associated virus Rep68 protein: toward a chromosome-targeting nonviral particle. J Virol. (1998) 72:7653-7658.

66. Pieroni L, Fipaldini C, Monciotti A, Cimini D*, et al*. Targeted integration of adeno-associated virus-derived plasmids in transfected human cells. Virology. (1998) 249:249-259.

67. Hacein-Bey-Abina S. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science. (2003) 302:415-419.

68. Gaspar H, Parsley K, Howe S, King D*, et al*. Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. The Lancet. (2004) 364:2181-2187.

69. Aiuti A, Cattaneo F, Galimberti S, Benninghoff U*, et al*. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med. (2009) 360:447-458.

70. Jacobson SG, Cideciyan AV, Ratnakaram R, Heon E*, et al*. Gene therapy for leber congenital amaurosis caused by RPE65 mutations: safety and efficacy in 15 children and adults followed up to 3 years. Arch Ophthalmol. (2012) 130:9-24.

71. LeWitt PA, Rezai AR, Leehey MA, Ojemann SG*, et al*. AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial. Lancet Neurol. (2011) 10:309-319.

72. Raper SE, Chirmule N, Lee FS, Wivel NA*, et al*. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab. (2003) 80:148-158.

73. Frank K, Hogarth D, Miller J, Mandal S*, et al*. Investigation of the cause of death in a gene-therapy trial. N Engl J Med. (2009) 361:161-169.

1.2. Biosafety considerations of gene therapy

The potential for genotoxicity in gene therapy is not unexpected. Initial studies investigating the integration site preferences of different viral vectors such as HIV, ASLV and MLV gammaretrovirus, drew attention to the potential for insertional mutagenesis arising from random or quasi-random genomic integrations, aggravated by the marked propensity of these vectors to target transcription start sites and active genes[1]. Even before reports of adverse events surfaced in clinical trials, a 2002 *in vivo* retroviral gene marking study of murine bone marrow cells already reported a high frequency of vector-induced hematopoietic disorders, including leukemia, caused in part by insertional activation of an oncogene[2]. Different strategies are therefore being actively explored to reduce the genotoxic potential of current viral vectors, mainly focused on devising methods for: (a) appropriate tissue targeting of systemically delivered vectors; (b) disabling the capacity for generating replication-competent viruses; (c) mitigating immune responses to vectors and/or transgene products; (d) avoiding germline modifications; (e) preventing unintended vector dissemination; and (f) directing the integration of transgenes into genomic safe harbors.

1.2.1. Immune response

Immune responses against systemically delivered vectors or transgene products expressed by gene-modified cells may affect the efficacy of gene therapy and, in some instances, culminate in serious life threatening inflammatory response[3] or other health hazards. Adaptive immunity gained from life-long exposure to natural viruses may also reduce the efficacy of certain viral vectors due to the presence of pre-existing antibodies against common viral antigens[4, 5]. Adenoviruses are one of the more immunogenic viral vectors, given their propensity to invoke a repertoire of different immune responses, including cytotoxic T-lymphocyte (CTL) response against viral components or transgene products, antibody mediated humoral response and cytokine-mediated inflammatory responses towards viral capsid or other viral components[4]. Potent immune responses may eliminate transgene expressing cells and reduce efficacy of treatment or, in the worst case scenario, induce an acute inflammatory response and cytokine storm as was the case in the first reported clinical gene therapy death[3]. The development of helper-dependent and gutless adenoviral vectors which are devoid of most viral genes has helped diminish their immunogenicity to some extent; more specifically, the adaptive immune response towards these vectors[6].

HSV vectors also induce cytopathic and inflammatory responses in human subjects[7]. Most other viral vectors, such as AAV, lentiviruses and retroviruses, are relatively less immunogenic compared to the aforementioned vectors. Recombinant AAV (rAAV) which are devoid of most viral genes may, however, invoke anti-AAV neutralizing antibodies against their capsid proteins which will prevent the possibility of re-administration of the same vectors if ever required[5]. Systemic delivery of naked plasmid DNA is also known to induce immune responses (due to antigenic unmethylated CpG motifs)[8], albeit of milder severity than most viral vectors. This risk can theoretically be mitigated by modifying the administration protocols or by using CpG-free plasmids[9]. In general, implantation of cells after e*x vivo* gene modification can be expected to be associated with lower risks of invoking immune responses compared with direct *in vivo* gene delivery. The issue of immunogenicity directed against transgene products[10] is a problem common to both viral and non-viral gene therapy that remains to be resolved through immunomodulation and immunotolerance induction strategies.

1.2.2. Insertional mutagenesis/ oncogenesis

Insertional mutagenesis is the induction of deleterious mutations to genes, promoters, enhancers or other regulatory elements that alter gene expression qualitatively or quantitatively as a consequence of exogenous vector integration into the genome. Although a major concern of integrating vectors, even non-integrating vectors have a low but finite possibility of random genomic integration[11].

Prior to cases of gene therapy-induced oncogenesis in recent clinical trials (section 1.1.4), the risk of malignant transformation from integrating vectors was considered theoretically plausible but unlikely to occur in practice. With hindsight, treatment-induced malignancies could have been predicted on the basis that as many as 1% of genes encoded in the genome are implicated in one or more forms of cancer[12]. Although oncogenesis is a process that requires multiple genetic hits, random integration of vectors into multiple genomic sites could be sufficient to generate the right "cocktail" of aberrations in different oncogenes and/or tumor suppressor genes[13]. Moreover, as the formerly regarded gene deserts are now known to be richly populated with different classes of non-protein-coding RNAs with key roles in cellular maintenance and cancer development[14, 15], evaluation of the genotoxic risk of integration events requires extra caution.

Commonly used viral vectors do not integrate randomly but have a propensity for transcriptionally active units and transcription start sites in mammalian

cells[1, 16]. Studies of integration profiles have been instrumental in developing integration maps of different viruses from which the potential genotoxicity of each type of viral vector may be assessed. Aside from the known cancer genes, disruptive integrations into other genes such as those necessary for cell survival or metabolism may be deleterious. Thus, insertional mutagenesis is a real risk that needs to be seriously addressed rather than dismissed as an inconsequential concern, as was the attitude prior to reports of adverse outcomes in gene therapy clinical trials.

Much has been learned about the molecular pathogenesis of oncogenesis associated with integrating viral vectors. MLV gammaretroviral vectors have a predilection for integrating close to transcription start sites[1] and to perturb their expression possibly due to the strong enhancer effect inherent in the LTRs[17]. However, this alone may not be sufficient for complete oncogenic evolution as ten patients treated with a similar MLV retroviral vector in a clinical trial for SCID-ADA did not develop untoward outcomes (median follow-up of 4 years)[18]. This has led to the speculation that other factors such as the nature of the expressed transgene (*IL2Rγ* versus *ADA*), the underlying disease, cell types selected for transgenic modification and other patient-specific intrinsic factors are necessary accessory factors to oncogenesis[19].

In contrast to retroviral vectors, no overt adverse events have been reported thus far from the use of other viral vectors such as lentiviral, adenoviral, HSV or AAV vectors. Some studies even suggest that lentiviral vectors pose significantly lower risks of insertional oncogenesis compared to retroviral vectors due to differences in their integration preferences[20]. Generally, non-integrating vectors such as adenoviruses, rAAV and HSV which are predominantly maintained as episomes are not considered to be mutagenic given their minute possibility of inducing rare random integrations in the genome. On the other hand, rAAV vectors which are largely maintained as episomes but are also known to integrate into the genome at low frequencies[21], must be considered as having intermediate risks.

Thus in summary, integrating vectors do bear potential risks of inadvertently affecting the genome. These can be in the form of deletions or large insertions in multiple genomic regions, dysregulation of endogenous genes, epigenetic effects and/or abnormal chromosome structures. Therefore, it is imperative that gene therapy modalities are comprehensively evaluated for potential genotoxicity. Therapeutic approaches that adopt *ex vivo* cell modification benefit from the advantage that biosafety assessments can be performed before *in vivo* implantation. Such biosafety

evaluations can be expected to reduce iatrogenic complications as treatment can be halted if significant risks are identified beforehand.

1.2.3. Germline transduction

An inherent disadvantage of *in vivo* gene delivery (especially with viral vectors) is the unintended dissemination of vectors to several tissues and organs with consequent transduction of several non-target cell types that could facilitate undesirable immune responses and genotoxicity. Generally, the main intent of gene delivery is to transduce a particular target population of cells relevant to the disease being treated, but more often than not this objective is difficult to realize with systemic delivery of vectors. Moreover, inadvertent transduction of immune cells, such as antigen-presenting cells, may provoke inflammatory responses[22] as well as invoke immune responses towards the transgene products and resulting in elimination of vectors and/or the transgene expressing cells[23]. The use of tissue/cell-specific promoters may reduce transgene expression in unintended cells to some extent and at least partially circumvent immunogenic effects[24]. Furthermore, regional administration of vectors to target organs may usefully restrict delivery to tissues which are intended for transgene expression. Modifications to vector surfaces to recognize and transduce specific cell types have also proved useful in narrowing the range of cell types or tissues transduced by systemically delivered viral vectors[24].

Yet another serious concern of systemic *in vivo* vector delivery is risk of inadvertent transmission of the transgene and vector-induced genome modifications through the germline. This could, in turn, pose mutagenic risks to future progeny. Germline transmission became a concern in a clinical trial for hemophilia B when semen samples of six of seven studied subjects who received systemic AAV2 vectors proved to be positive for viral vector sequences. This unintended shedding of viral vector components and dissemination to gonadal tissues raised concerns of spermatogonial transduction and potential vertical transmission of viruses[25]. Long-term follow-up allayed these fears as the presence of viral vectors in semen was found to be transient. Further studies in rabbits showed that AAV vectors could be present in seminal fluid without effectively transducing sperm cells[26], i.e. vector was not detected in cellular fractions. Nonetheless, this episode serves as a caution that evaluating the risks of germline transmission is essential, especially when viruses are delivered systemically.

1.2.4. Tools for evaluating potential for genotoxicity

The reality of vector-induced oncogenesis need not be a fatal impediment to the goal of clinical gene therapy if gene transfer approaches are rigorously evaluated for potential genotoxicity. Tools are now available to interrogate transgenically-modified cells *ex vivo* for genomic alterations and to evaluate their tumorigenic potential. The ability to perform comprehensive biosafety assessments *ex vivo* before proceeding to *in vivo* treatment is feasible and opens a way to exploit the benefits of gene replacement while minimizing treatment risks to a clinically acceptable level.

A first step to genotoxicity analysis of any given modality is to review databases for adverse outcomes encountered in past or ongoing clinical trials which can be accessed at several websites e.g. Wiley clinical trials database (http://www.wiley.com/legacy/wileychi/genmed/clinical/), the US National Institutes of Health ClinicalTrials.gov (http://www.genetherapynet.com/clinicaltrialsgov.html) and Clinigene (http://www.clinigene.eu/search-published-human-gene-therapy-clinical-trials-database/).

The preceding section summarized complications that may arise from gene therapy. This section now focuses on the biosafety assessment of *ex vivo* gene modified cells, with an emphasis on key features to monitor and molecular biology tools that aid the evaluation. The importance of bioinformatic tools in biosafety evaluation cannot be overemphasized. This section will also highlight useful programs, internet resources and databases.

1.2.4.1. Mapping genome integration sites

It is imperative to document integration events in gene modified cells, and prudent to do so even for episomal vectors that have a low probability of random integration[11, 27]. Integration events are detailed with reference to their physical distance relative to promoter sites, transcription start sites, exons or introns, oncogenes, tumor suppressor genes, non-protein coding genes, CpG islands, repetitive elements and transcription factor and micro-RNA binding sites. Such integration profiles aid genotoxicity risk evaluation when comparing across vector types, modified cell types and the nature of transgenes.
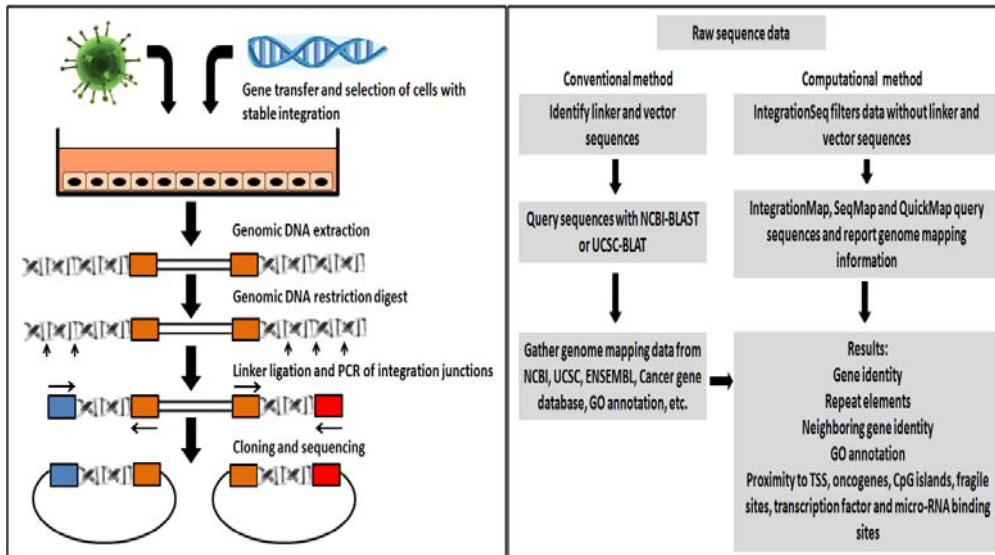
Figure 1.2.1   Experimental recovery of integration events and computational analysis of integration site distribution in mammalian cell genomes. Left: Integration events in cells are retrieved by digesting genomic DNA with restriction enzymes that do not cleave within the vector sequences. Appropriate adapters are ligated to restriction fragments to serve as priming sites for PCR amplification of integration junctions which can be cloned and sequenced. (*Adapted from Ciuffi A, et al., 2009*)[28]. **Right:** Vector flanking raw sequence data may be selected with programs such as IntegrationSeq and queried using UCSC-BLAT or NCBI-BLAST to retrieve relevant genomic information. Computational programs such as IntegrationMap, SeqMap and QuickMap automate the process of genome mapping and provide the necessary annotations for biosafety assessment. (*Adapted from Peters B, et al., 2008*)[29].

Integration events within cells can be experimentally retrieved and identified by plasmid rescue, ligation mediated PCR (LM-PCR)[30], inverse PCR[31] or linear amplification mediated PCR (LAM-PCR)[32]. Sequence data can be analyzed for vector-flanking sequences by programs such as IntegrationSeq[33] which may then be queried in genome database programs such as NCBI-BLAST (http://blast.ncbi.nlm.nih.gov/) or UCSC-BLAT (http://ww.genome.ucsc.edu/) to identify their genomic positions (**figure 1.2.1**). In recent years, several programs have been developed to automate the process of genome mapping. IntegrationSeq[33], SeqMap[29] and QuickMap[34] are examples of web-based programs that are useful for annotating genome mapping information such as proximity to genes, neighboring gene identity, exon/intron localization, distance from transcription start sites, repeat element localization and Gene Ontology functions. Recently developed QuickMap (http://www.gtsg.org) provides a more comprehensive evaluation that includes information about proximity to oncogenes, pseudogenes, CpG islands, fragile sites, transcription factor and micro-RNA binding sites. Identity of potential cancer genes can be derived from lists compiled from the human cancer gene census[12] or the

retrovirus and transposon tagged cancer gene database, RTCGD (http://variation.osu.edu/rtcgd/about_us.html). Another useful database with a comprehensive compilation of known oncogenes and tumor suppressor genes[35] is hosted by the University of Pennsylvania School of Medicine (http://www.bushmanlab.org/links/genelists).

Another use of profiling of genomic integration sites is for long-term monitoring of the clonality of *in vivo* implanted gene modified cells[36]. Integration profiles determined pre-implantation can be periodically monitored post-implantation to detect the emergence of dominant clones. Deviation from a polyclonal pattern of growth could imply selection of a dominant clone by virtue of a survival advantage or a greatly increased proliferation rate. This ought to trigger close scrutiny for the likelihood of insertional oncogenesis. Gerrits *et al*. have recently developed tagged vectors with variable unique barcode signatures for tracking different clones *in vivo*[37]. Such innovative techniques could be applied to enhance monitoring the clonality of implanted cells *in vivo* and increase the sensitivity of detecting potential oncogenic alterations.

1.2.4.2. Characterizing the modified genome

There is a sound basis to expect that integrating and non-integrating vectors may alter genomic architecture. Copy number gains and deletions have been observed in transformed cancer cell lines and to a lesser extent in cells modified with gene transfer vectors[27]. Recent advances in array-based methods have made it possible to study genome-wide amplifications or deletions at high resolution with probes that tile the genome with an average inter-probe interval of 2.5 kb[38]. As with most array-based techniques, copy number analysis relies on a relatively homogeneous population of cells as alterations in a minor population within a polyclonal population may be masked or underrepresented in the analysis that would otherwise only reveal the copy number profile of the dominant cell population.

Another type of potentially pathogenic genomic alteration of concern are numerical or structural chromosomal abnormalities, i.e. aneuploidy and/or structural abnormalities such as deletions, translocations and inversions, which are common hallmarks of transformed cells. Several studies have reported rare cytogenetic abnormalities in cells treated with rAAV[39], retroviral vectors[40] and non-viral vectors such as phiC31 phage integrase-mediated plasmid integration[41, 42]. The inciting causes of such cytogenetic abnormalities are unclear, namely whether from direct effects of vector integration and repair or from recombination events secondary to vector

integration. Also unclear is whether these rare cytogenetic changes are related to gene transfer manipulations or simply reflect the intrinsic low frequency of chromosomal anomalies that occur even in normal somatic cell populations[43].

Gross chromosomal rearrangements in gene modified cells can be evaluated by spectral karyotyping (SKY) or multi-color fluorescence *in situ* hybridization (FISH). Karyotyping requires examination of a sufficient number of good quality metaphase chromosomes if rare rearrangements are not to be missed. Array-based comparative genomic hybridization detects copy number abnormalities (deletions or amplifications) at high resolution, provided a fairly homogeneous cell population is analyzed. However, even high resolution copy number analysis will not detect aberrations in a rare subpopulation of cells. Genome sequencing to identify vector integration junctions can potentially identify translocations at the highest (single nucleotide level) resolution provided junctional fragments can be confidently identified. However, this method (currently performed at relatively high, albeit decreasing, cost) generates large datasets that require specialized bioinformatic analysis and awareness of technical artifacts[44]. Therefore effective cytogenetic analysis should be combined with sequencing techniques (for integration site retrieval), multi-color karyotyping, whole genome copy number profiling and possibly deep whole genome sequencing analysis as a complementary suite of techniques to completely characterize the genome of gene modified cells.

1.2.4.3. Transcriptome and epigenome analysis

A necessary complement to genome analyses is to determine effects of gene transfer (however accomplished, but especially if the transgene is known to have integrated) on the transcriptome of gene modified cells. In this regard, it is worth noting that vector insertions are often accompanied by deletions of genomic regions[45] that may in turn alter the epigenetic status of the cell if key histone proteins, histone modifying and DNA methylating enzymes are affected. Thus it may also be relevant to determine effects of gene transfer on the epigenome.

Comparing the global transcriptomes of naïve and vector treated cells may help to identify genes whose expressions are perturbed. Many technical platforms based on hybridization to gene-specific oligonucleotide probes and RNA-Seq[46] have been developed for genome-wide transcriptome analysis and, being unbiased, are the methods of choice. Such data, in practice, reveal significantly altered gene expression mainly in the dominant cell population, though not necessarily in minor subpopulations. Characterizing the transcriptome of a homogeneous or, preferably,

clonal population of cells with a single known vector integration is ideal as it is more informative. The presence of multiple integration sites in a clonal population confounds attempts to distinguish effects attributable to any particular integration. Likewise, the study of a heterogeneous cell population could mask the transcriptional features of a minor subpopulation. To set stringent standards for biosafety assessment, microarray studies are useful only when a sufficient number of clonal populations from different integration sites are characterized. Given that viral vectors mediate integrations into multiple sites, such clonal studies are highly impractical. Clonal studies are especially important when integrations occur close to oncogenes and tumor suppressors. Transcriptome analysis aims not only to identify individual genes with significantly altered expression but should also map individual aberrations to molecular pathways. There is a plethora of non-proprietary microarray analysis and bioinformatic software tools for data evaluation and analysis. For example, useful tools are hosted by Gene Ontology (http://www.geneontology.org/GO.tools.microarray.shtml), Genomics and Bioinformatics Group from NIH (http://discover.nci.nih.gov/tools.jsp) and Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/home.jsp).

Epigenetic changes refer to alterations in the acetylation, methylation, sumoylation and phosphorylation patterns of histone proteins, which in turn may affect the dynamic chromatin architecture and determine the active or repressed status of genes. It also encompasses changes in CpG methylation near promoter regions which may influence gene expression. Transgene and vector integrations may directly attenuate gene expression or have negative or positive effects on genes based on their effects on histone modifying or DNA methylating enzymes. Global characterization of epigenomes presently combine global transcriptome analysis, cytosine methylation patterns, nucleosome positioning assays and chromatin immunoprecipitation (ChIP)-based assays to determine transcription factor binding sites[47]. The on-going human epigenome project (http://www.epigenome.org/) that aims to document the DNA methylation patterns of all human genes is likely to provide invaluable insights into the role of epigenetics in human diseases. However, the study of epigenetics is currently hampered by a lack of simple, high quality and high-throughput techniques. Technical advances should deepen knowledge of this important domain of human genetics.

1.2.4.4. *In vitro* and *in vivo* tumorigenicity studies

Transformed cells acquire altered phenotypes that can be detected *in vitro*. Anchorage independent growth, loss of contact inhibition, resistance to apoptosis, increased proliferation rate and extended cell passaging are common characteristics of transformed cells.

Simple *in vitro* assays demonstrate anchorage independent growth and increased proliferation rates of cells. The soft agar colony formation assay involves enumerating colonies (clonal propagation of cells) formed from individual cells in the absence of substrate adhesion. Anchorage independent cells typically form colonies while normal cells do not as they rely on surface attachment for proliferation. Assays that quantify incorporation of bromo-deoxyuridine (BrdU), reduction of tetrazolium compounds (e.g. 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) and colony formation are direct or indirect measures of cellular proliferation rates. Modlich and colleagues[48] recently introduced the *in vitro* immortalization assay which tests tumorigenic potential of virally transduced murine HSCs based on their replating capacity, thus obviating the need for *in vivo* testing in animals.

Most *in vitro* biosafety assays seek to evaluate deviations from normal cellular characteristics. A more clinically relevant evaluation of tumorigenicity would be to determine the potential to induce tumors *in vivo*. Two main models are used for this purpose. In the first model, gene modified human cells are implanted into immunocompromised mice that are known to support the engraftment of xenogeneic cells. It is helpful to know that different strains of immunocompromised mice have different capacities to mount immune responses depending on which components of the immune system are still functional. Mouse strains that are most severely immunocompromised can be expected to have high sensitivity as tumorigenic hosts because even low numbers of implanted cells could give rise to visible tumors. Such sensitive models are useful for the detection of rare populations of oncogenic cells in a heterogeneous population of otherwise untransformed cells. The absence of tumor formation should not immediately exonerate cells of their tumorigenic potential. It is essential to establish from immunohistology of the implantation sites or in the case of HSCs implantation, immunocytometric blood analysis that the implanted cells have indeed engrafted *in vivo* in animals that fail to form tumors. The second model is useful to evaluate the genotoxic potential of HSCs transduced with different gene therapy vectors. It is based on the transduction and transplantation of HSCs derived from a tumor-prone mouse model that lacks the tumor suppressor, cyclin dependent kinase inhibitor 2A *(cdkn2a)* gene[49]. This assay thus evaluates tumorigenic risk in an already tumor-prone cell line and has been used to compare the oncogenic potential

of retroviral and lentiviral vectors, and to assess the benefits of incorporating SIN-LTRs in these vectors. However, a caveat is that owing to the intrinsic oncogenic potential of the *cdkn2*-/- HSCs, the effects of subtle but relevant insertional mutagenic events may be masked or misinterpreted. Besides murine models, long-term studies can also be performed in pre-clinical animals such dogs and non-human primates[50] where the clonality of implanted cells can be dynamically monitored by documenting integration profiles of recovered cells to ascertain if dominant clones with clone-specific integration patterns have emerged.

1.2.5. References

1. Mitchell RS, Beitzel BF, Schroder ARW, Shinn P*, et al*. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. (1). PLoS Biol. (2004) 2:E234.

2. Li Z, Dullmann J, Schiedlmeier B, Schmidt M*, et al*. Murine leukemia induced by retroviral gene marking. Science. (2002) 296:497.

3. Raper SE, Chirmule N, Lee FS, Wivel NA*, et al*. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab. (2003) 80:148-158.

4. Chuah MKL, Collen D, VandenDriessche T. Biosafety of adenoviral vectors. Curr Gene Ther. (2003) 3:527-543.

5. Mingozzi F, Maus MV, Sabatino DE, Murphy SL, *et al*. CD8(+) T-cell responses to adeno-associated virus capsid in humans. Nat Med. (2007) 4:419-422.

6. Segura M, Alba R, Bosch A, Chillon M*, et al*. Advances in helper-dependent adenoviral vector research. Curr Gene Ther. (2008) 8:222-235.

7. Gogev S, Schynts F, Meurens F, Bourgot I*, et al*. Biosafety of herpesvirus vectors. Curr Gene Ther. (2003) 3:597-611.

8. Weiner GJ. The immunobiology and clinical potential of immunostimulatory CpG oligodeoxynucleotides. J Leukoc Biol. (2000) 68:455-463.

9. Niidome T, Huang L. Gene therapy progress and prospects: nonviral vectors. Gene Ther. (2002) 9:1647-1652.

10. Tripathy SK, Black HB, Goldwasser E, Leiden JM*, et al*. Immune responses to transgene-encoded proteins limit the stability of gene expression after injection of replication-defective adenovirus vectors. Nat Med. (1996) 2:545-550.

11. Wang Z, Troilo PJ, Wang X, Griffiths TG, *et al*. Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. Gene Ther. (2004) 11:711-721.

12. Futreal PA, Coin L, Marshall M, Down T, *et al*. A census of human cancer genes. Nat Rev Cancer. (2004) 4:177-183.

13. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. (2000) 100:57-70.

14. Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res. (2011) 71:3-7.

15. Farazi TA, Spitzer JI, Morozov P, Tuschl T, *et al*. miRNAs in human cancer. J Pathol. (2011) 223:102-115.

16. Beard BC, Dickerson D, Beebe K, Gooch C, *et al*. Comparison of HIV-derived lentiviral and MLV-based gammaretroviral vector integration sites in primate repopulating cells. Mol Ther. (2007) 15:1356-1365.

17. Modlich U, Bohne J, Schmidt M, von Kalle C, *et al*. Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. Blood. (2006) 108:2545-2553.

18. Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, *et al*. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med. (2009) 360:447-458.

19. Baum C. Insertional mutagenesis in gene therapy and stem cell biology. Curr Opin Hematol. (2007) 14:337-342.

20. Montini E, Cesana D, Schmidt M, Sanvito F, *et al*. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol. (2006) 24:687-696.

21. Nowrouzi A, Penaud-Budloo M, Kaeppel C, Appelt U, *et al*. Integration frequency and intermolecular recombination of rAAV vectors in non-human primate skeletal muscle and liver. Mol Ther. (2012) 20:1177-1186.

22. Schnell MA, Zhang Y, Tazelaar J, Gao GP*, et al*. Activation of innate immunity in nonhuman primates following intraportal administration of adenoviral vectors. Mol Ther. (2001) 3:708-722.

23. Brown BD, Lillicrap D. Dangerous liaisons: the role of "danger" signals in the immune response to gene therapy. Blood. (2002) 100:1133-1140.

24. Frecha C, Szécsi J, Cosset FL, Verhoeyen E*, et al*. Strategies for targeting lentiviral vectors. Curr Gene Ther. (2008) 8:449-460.

25. Manno CS, Arruda VR, Pierce GF, Glader B*, et al*. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med. (2006) 12:342-347.

26. Schuettrumpf J, Liu JH, Couto LB, Addya K*, et al*. Inadvertent germline transmission of AAV2 vector: findings in a rabbit model correlate with those in a human clinical trial. Mol Ther. (2006) 13:1064-1073.

27. Stephen SL, Sivanandam VG, Kochanek S. Homologous and heterologous recombination between adenovirus vector DNA and chromosomal DNA. J Gene Med. (2008) 10:1176-1189.

28. Ciuffi A, Ronen K, Brady T, Malani N*, et al*. Methods for integration site distribution analyses in animal cell genomes. Methods. (2009) 47:261-268.

29. Peters B, Dirscherl S, Dantzer J, Nowacki J*, et al*. Automated analysis of viral integration sites in gene therapy research using the SeqMap web resource. Gene Ther. (2008) 15:1294-1298.

30. Laufs S, Gentner B, Nagy KZ, Jauch A*, et al*. Retroviral vector integration occurs in preferred genomic targets of human bone marrow-repopulating cells. Blood. (2003) 101:2191-2198.

31. Silver J, Keerikatte V. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J Virol. (1989) 63:1924-1928.

32. Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K*, et al*. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods. (2007) 4:1051-1057.

33. Giordano FA, Hotz-Wagenblatt A, Lauterborn D, Appelt JU*, et al*. New bioinformatic strategies to rapidly characterize retroviral integration sites of gene therapy vectors. Methods Inf Med. (2007) 46:542-547.

34. Appelt JU, Giordano FA, Ecker M, Roeder I*, et al*. QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. Gene Ther. (2009) 16:885-893.

35. Wang GP, Garrigue A, Ciuffi A, Ronen K*, et al*. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res. (2008) 36:e49.

36. Wang GP, Berry CC, Malani N, Leboulch P*, et al*. Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial. Blood. (2010) 115:4356-4366.

37. Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K*, et al*. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood. (2010) 115:2610-2618.

38. Hester SD, Reid L, Nowak N, Jones WD*, et al*. Comparison of comparative genomic hybridization technologies across microarray platforms. J Biomol Tech. (2009) 20:135-151.

39. Miller DG, Trobridge GD, Petek LM, Jacobs MA*, et al*. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J Virol. (2005) 79:11434-11442.

40. Modlich U. Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. Blood. (2005) 105:4235-4246.

41. Liu J, Jeppesen I, Nielsen K, Jensen TG, *et al*. Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther. (2006) 13:1188-1190.

42. Ehrhardt A, Engler JA, Xu H, Cherry AM, *et al*. Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther. (2006) 17:1077-1094.

43. Varella-Garcia M, Chen L, Powell RL, Hirsch FR, *et al*. Spectral karyotyping detects chromosome damage in bronchial cells of smokers and patients with cancer. Am J Respir Crit Care Med. (2007) 176:505-512.

44. Koboldt DC, Ding L, Mardis ER, Wilson RK, *et al*. Challenges of sequencing human genomes. Brief Bioinform. (2010) 11:484-498.

45. Miller DG, Rutledge EA, Russell DW. Chromosomal effects of adeno-associated virus vector integration. Nat Genet. (2002) 30:147-148.

46. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. (2009) 10:57-63.

47. Fazzari MJ, Greally JM. Introduction to epigenomics and epigenome-wide analysis. Methods Mol Biol. (2010) 620:243-265.

48. Modlich U, Navarro S, Zychlinski D, Maetzig T, *et al*. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. (1). Mol Ther. (2009) 17:1919-1928.

49. Montini E, Cesana D, Schmidt M, Sanvito F, *et al*. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest. (2009) 119:964-975.

50. Kim YJ, Kim YS, Larochelle A, Renaud G, *et al*. Sustained high-level polyclonal hematopoietic marking and transgene expression 4 years after autologous transplantation of rhesus macaques with SIV lentiviral vector-transduced CD34+ cells. Blood. (2009) 113:5434-5443.

1.3. Recent enhancements in biosafety of gene therapy

Comprehensive molecular studies that were undertaken in light of adverse outcomes in clinical trials of gene therapy have advanced our understanding of their likely mechanisms[1]. This has, in turn, spurred the development of safer approaches. In parallel, more sensitive genome-wide techniques now available for biosafety evaluations enable higher confidence in pre-clinical assessments before treatments are adopted in clinical trials. This section reviews recent developments in gene transfer by non-viral vectors that could potentially enhance biosafety.

1.3.1. Improvements to integrating non-viral vectors

Non-viral vectors have found useful applications largely in the laboratory and pre-clinical settings but represent only 24% of all vectors used in clinical gene therapy trials. The fact that viruses have evolved over millennia to become effective infectious agents understandably makes them superior in many aspects as gene transfer agents. The ultimate goal of designing synthetic non-viral vectors is to combine the positive traits of viruses without the negative threats of genotoxicity. Significant improvements have been made to methods of non-viral vector delivery[2] with reported efficiencies that rival those achieved with viral transductions. Two classes of non-viral vectors may contribute to improved biosafety of gene therapy, namely episomally maintained vectors and integrating vectors with safer integration profiles. In the context of treating diseases caused by a single gene defect, the ultimate goal of an ideal gene transfer vector is to deliver durable and appropriately regulated transgene expression either from an autonomously replicating artificial chromosome, episomal plasmid or from transgenes integrated into safe genomic harbors. This section reviews recent progress in developing non-viral integrating vectors with safer integration profiles.

1.3.1.1. Transposase, recombinase and integrase

Transposases and recombinases are two classes of site-specific genome modifying agents. These enzymes recognize and bind to short stretches of DNA sequences within the vector and in the genome to mediate the integration of exogenous vector DNA into the genome. Analysis of the integration spectrum of several transposases and recombinases identified some that mediate quasi-random or sequence specific integrations into the genome, a distinct advantage over randomly integrating viral vectors. Transposases and recombinases are also less immunogenic[3], transposons have lower inherent enhancer/promoter activity on neighboring genes[4]

and induce fewer epigenetic effects at genomic integration sites[5], relative to viral vectors. Given their activity in mammalian cells, these non-viral integrating systems evoke exciting possibilities for development into safer alternatives than randomly integrating vector systems. Several different classes and strains of transposases and recombinases have been discovered and studied as gene therapy agents. An important concern is their relatively relaxed stringency of site-specific integrations which again raises the spectre of insertional mutagenesis. Therefore, a major effort has been directed at improving specificity. Another cautionary note is the low risk of unintended integration of the transposase or recombinase, which could have deleterious effects on the genome. Such risks may be minimized or abrogated by using messenger RNA (mRNA) rather than DNA to deliver the recombineering proteins. The following sections highlight advances in the more commonly used transposases and recombinases.

1.3.1.1.1.   Transposases – Sleeping Beauty, PiggyBac and *Tol2*

The Sleeping Beauty (SB) system was reconstructed from a molecular fossil in fish genomes. It belongs to the *tc1/mariner* superfamily and is one of the most widely investigated transposase systems to date. SB transposase mediates genomic integration of sequences flanked by an inverted tandem repeat (ITR) at each transposon end, preferentially into TA dinucleotides located within genomic regions with increased local bendability[6], via a "cut-and-paste" mechanism. Integrations are quasi-random, without any preference for transcriptionally active regions[7]. Optimized SB has a transposition efficiency[8] of 2.5 - 17%. Stable integration using this system has enabled long-term transgene expression in a variety of mammalian cells and animal models[9, 10]. Owing to the randomness of integrations, SB systems have been used also as tools for discovery of new oncogenes both in *in vitro* and *in vivo* models[11]. It is worth reiterating that these SB systems are different from those used in gene therapy applications. SB systems used in oncogene discovery are deliberately modified *via* incorporation of strong transcriptional enhancers and splice acceptor sites to be potently oncogenic[11]. Thus far, the use of SB as a gene therapy agent in animal models has not been associated with any evidence of tumorigenesis[12]. Inherent limitations of the SB system include limited cloning capacity, inhibition of transposition at high transposase concentrations and lack of targeting specificity of integrations.

SB systems are prone to reduced transposition with increased cargo load. Zayed *et al*. demonstrated that it was possible to retain transposition with inserts

greater than 10 kb by using a sandwiched vector in which the gene of interest was flanked by two complete SB elements in inverted orientations[13].

Initial studies with naive SB system revealed their inherently low transposition efficiencies. Many modifications have since been introduced to create hyperactive versions of SB with increased transposition activity such as SB10[14], SB11[15] and SB100X[16]. The hyperactive SB100X, which was reported to have a 100-fold increased transposition activity, was discovered by high-throughput screening of mutants created by a PCR-based DNA shuffling strategy. Using these improved versions of SB, efficient transposition has been reported in a variety of human primary cells such as cord blood derived CD34+ hematopoietic progenitor cells[17], primary T cells[7] and embryonic stem (ES) cells [18].

The issue of non-specific targeting by SB has been another prime focus of research aimed at inducing site-specific integration. An ideal modification would enable SB to direct transposition to a single pre-defined "safe harbor" in the genome. Skewing the random integration pattern of SB towards a more targeted profile would be hailed as an improvement. Several groups have attempted to do this by incorporating specific DNA-binding domains (DBD) either to the SB transposase[19] (**Figure 1.3.1; top panel**), the transposon bearing the gene of interest[20] or *via* a fused DBD-protein binding domain (PBD) that interacts with the transposase without modifying it[21]. The first strategy of fusing DBDs such as E2C (a synthetic zinc finger protein that recognizes an 18 bp target site in the 5'-untranslated region of the human *ERBB2* gene) or Gal-4 to the transposase met with limited success. In a second strategy, Ivics and collaborators demonstrated re-targeted integrations by incorporating a fusion of two DBDs to direct the transposon bearing the gene of interest to specific genomic sites where transposition could be mediated by the transposase[20] (**Figure 1.3.1; bottom panel**). A third strategy utilizing a fusion of peptides that interact with the genomic locus of choice (*via* DBD) and the transposase (*via* PBD) without compromising transposase activity has also been reported by the same group[20] (**Figure 1.3.1; middle panel**). However, it must be noted that none of these site targeting modifications has yet been successfully translated to human gene therapy applications, probably because of the relatively poor efficiencies of re-targeting specificity.
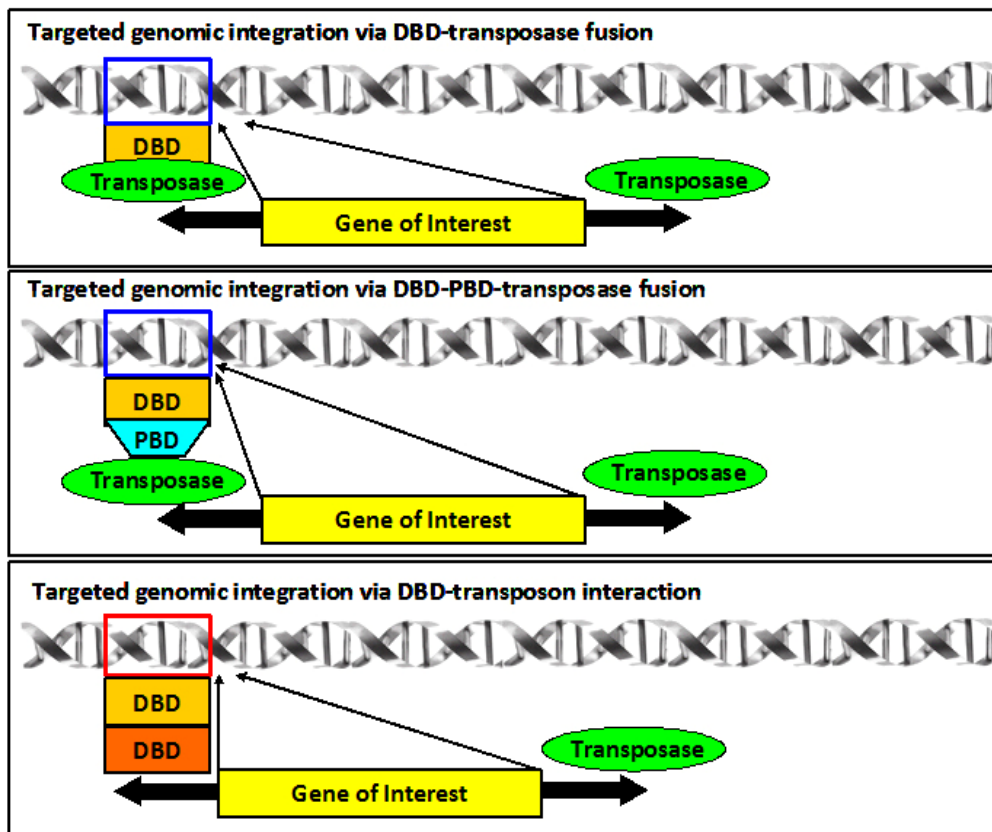
Figure 1.3.1   Re-directing targeting specificities of transposase/ transposons. Transposase can be re-targeted to a different specific genomic site by direct fusion with a (**top panel**) DNA-binding protein or (**middle panel**) indirectly via a protein-binding domain. Targeting can also be achieved with (**bottom panel**) a pair of DBD-fusion proteins recognizing specific sites in the genome and sequences within the transposon vector. (*Adapted from Izsvak Z, et al., 2010; ref. 62*).

The lack of propensity of non-viral integrating SB systems for active transcriptional units may make them safer than retroviral and lentiviral vectors. This has generated the idea of hybrid vectors that combine SB transposition with improved delivery by integrase-defective lentiviruses[22]. However, until there are effective solutions for improving the specificity of integrations, the SB system may have only limited appeal for clinical gene therapy. The only human clinical trial (phase I/II, NIH-OBA no. 0804–922) utilizing the SB system aims to redirect the specificity of T-cells by stable expression of CD19-specific chimeric antigen receptors mediated by the SB11 transposase system[23]. However, caution should be exercised before more transposon-based systems are translated to clinical applications, especially in light of the unexpectedly high copy number of random integrations of transposase plasmid in human primary T cells[24].

PiggyBac (PB) transposase, isolated from the cabbage looper moth (*Trichoplusia ni*), is another class of transposase which is active in human and murine

cells[25]. The PB system has been used to effectively reprogramme iPSCs[26] and mutagenize mice for cancer gene discovery[27]. PB demonstrated higher transposase activity than SB11 and could also be modified to incorporate DBD without loss of transposase activity[28]. Several improved versions of PB have been reported. Liang *et al.*[29] demonstrated increased chromosomal transposition with a codon optimized PB and, more recently, reported the development of a hyperactive PB with a 7-fold increase in integration activity and showed its application for generating murine iPSCs[30].

The *Tol2* transposon of the hobo/Activator/Tam3 (hAT) family of elements, derived from the medaka fish (*Orizyas latipes*), is active in human cells[31]. Like PB, *Tol2* also tolerates overproduction inhibition and, unlike the SB system, has a large cloning capacity (up to 18 kb). However, both PB and *Tol2* systems have significantly increased integrations into transcription start sites (TSS), CpG islands and DNaseI hypersensitive regions. Not surprisingly, transcriptional levels of neighboring genes close to integration sites in human T cells were altered[7]. This suggests a greater risk of insertional mutagenesis/oncogenesis compared with the SB system. In this respect, the PB and *Tol2* transposases may be better suited for applications where high frequencies of mutagenesis are desired, such as cancer gene discovery[27].

1.3.1.1.2. Cre-*loxP*/ Flp-Frt

Cre (cyclization recombinase) recombinase, a tyrosine recombinase of the bacteriophage P1 family, catalyses site-specific DNA recombination at a 34 bp site called *loxP*, by strand cleavage, exchange and ligation. It can mediate the integration of a *loxP*-bearing vector DNA into pre-integrated *loxP* sites or endogenous pseudo *loxP* sites in the genome of eukaryotic cells[32]. The Cre/*loxP* system has found wide applications in ES cell engineering and in developing knock-in or knock-out transgenic mice. There have been efforts to engineer improved versions of Cre recombinases to expand their applications. For example, a modified Cre recombinase recognizes a novel *loxH* site[33] in human chromosome 22. Other improved versions of Cre exhibit more efficient recombination[34]. Whilst the concept of Cre/*loxP*-mediated targeted gene integration appears ideal, certain biosafety concerns have impeded development of this system for clinical applications. These include evidence of genotoxicity and chromosomal translocations induced by high levels of Cre recombinase expression in mouse and mammalian cells, possibly due to recombination between endogenous cryptic *loxP* sites[35, 36]. Consequently, use of the

Cre/*loxP* system for genetic cell modifications remains confined to non-clinical applications.

The flippase (Flp) recombination enzyme, a tyrosine recombinase from *Saccharomyces cerevisiae*, mediates recombination of DNA fragments between two flippase recognition target (*FRT*) sites by mechanisms similar to the Cre/*loxP* system and has been utilized to integrate vector DNA sequences flanked by *FRT* sites into pre-integrated *FRT* sites within the genome in human cells. Thus far, there are no reports of targeted integration of *FRT*-containing sequences into non-native *FRT* sites in human cells. However, Flp recombinase can be mutated to recognize non-native sequences[37]. Flp with increased thermostability and greater recombination activity has also been developed[38] and applied in plant biotechnology.

Cre and Flp are examples of site-specific recombinases (SSR) frequently used in recombinase-mediated cassette exchange (RMCE).These two well characterized systems have frequently been used in forward genetics for directed integrations of exogenous DNA into specific chromosomal sites where target recombination elements (*loxP* or *FRT*) have been pre-integrated by homologous recombination (HR) or in reverse genetics for marker gene excision or deletion. RMCE strategies are helpful in deriving and investigating isogenic panels (all harboring integrations at the same genomic locus) of cell lines and ES cells without confounding positional effects, as would occur with random integrations[39]. This has led to commercial tools such as the FLP-In system (Invitrogen Corporation) for generating stably expressing cell-lines. Similarly, targeting strategies have been useful for generating "knock-in" or "knock-out" transgenic mice using germ-line cells modified with the Cre or Flp systems. RMCE has also been applied to create large cloning vectors such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs)[40], and to model human chromosomal translocations[41]. An inherent limitation of the Flp/Cre systems is their propensity to mediate reversible excision or inversion of integrated vectors from the genome owing to their continued expression. This has been partly overcome by using heterospecific mutant *loxP* or *FRT* sites[33, 37] and inverted orientations of recombination target sites[42]. In summary, the Cre and Flp systems have found greater applications in plant and animal biotechnology and transgenesis rather than as integrating agents for human gene therapy. The requirement for pre-integrated recombination target sites by homologous recombination into specific loci is not a trivial endeavor as it occurs at extremely low frequency (about 1 in 10^6 cells). Therefore deriving a sufficient number of *ex vivo* modified cells for gene therapy could be quite a challenge. The use of alternate SSRs

such as phiC31 integrase which mediate direct recombination into endogenous sites in genome or other agents that enhance integration of vector sequences by homologous recombination may therefore be more feasible approaches.

1.3.1.1.3. PhiC31 bacteriophage integrase

The phiC31 bacteriophage integrase is another class of SSR that has been quite thoroughly investigated for achieving therapeutic transgenesis. In its natural context, phiC31 integrase expressed by the bacteriophage of *Streptomyces lividans*, mediates integration of the bacteriophage genome into the bacterial genome *via* a recombination process between short stretches of DNA sequences known as phage attachment site (*attP*, 39 bp) and bacterial attachment site (*attB*, 34 bp). Several groups have shown that phiC31 integrase integrates plasmid vector sequences bearing *attB* sequences into either pre-integrated *attP* sites or pseudo *attP* sites naturally found in mammalian genomes[43] (**Figure 1.3.2**). Based on a central 28 bp sequence thought to be crucial for recombination, the similarity between an *attB* and a wild type *attP* sequence was approximated to be 50% while that between the wild type *attP* and pseudo *attP* consensus sequences, derived from sequence analysis of retrieved phiC31 integrase-mediated junctions, was about 64% [44] (**Figure 1.3.3**).
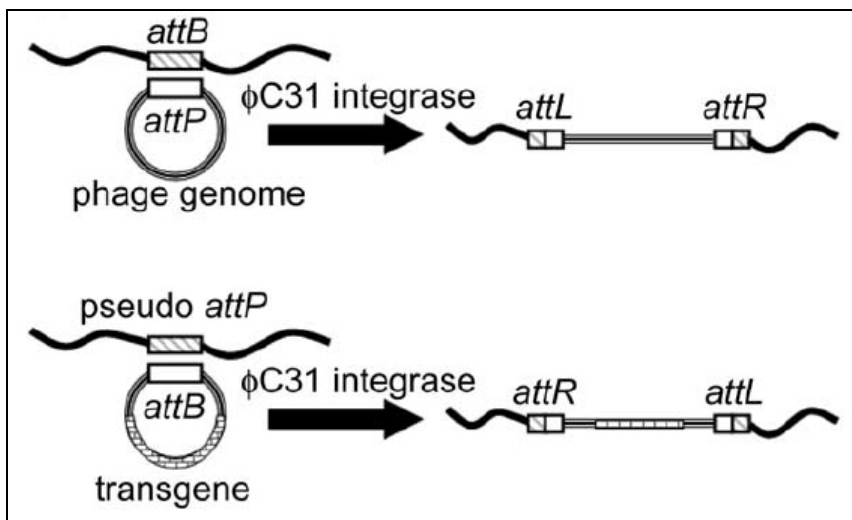
Figure 1.3.2   PhiC31integrase-mediated integration. (**Top**): Natural infection mechanism of *Streptomyces lividans* resulting in integration of bacteriophage genome into bacterial host genome following phiC31 integrase-mediated recombination between *attP* and *attB* sites in bacteriophage genome and bacterial chromosomes, respectively. (**Bottom**): PhiC31 integrase mediated transgene integration following recombination between *attB*-bearing plasmid DNA and pseudo *attP* sites in mammalian genomes. In both cases, integration results in an irreversible insertion of vector sequences flanked by *attL* and *attR* half-sequences which are refractory to further recombination. *(Figure taken from ref.45)*

PhiC31 integrase, a 68 kDA protein encoded by 613 amino acids, is a serine recombinase by virtue of serine residues within the catalytic domains that are essential for the recombination process. The protein consists of an N-terminal catalytic domain responsible for recombination and a C-terminal DNA-binding domain. PhiC31 integrase mediates site-specific recombination by initially binding to and forming a synaptic complex between recognition sequences, followed by a cut-and paste mechanism to mediate unidirectional integration of an *attB*-bearing vector sequence to *attP* or pseudo *attP* sequences in mammalian genomes, without requiring any additional host factors. Genomic integration results in an irreversible insertion of vector sequences flanked by *attL* and *attR* sequences which are refractory to further recombination by the integrase, unlike the reversible Cre/Flp systems. Analysis of phiC31 integrase-mediated integration events revealed that cross-over or integration usually occurs over a central TTG sequence within the *attP* consensus sequence[44] (**Figure 1.3.3**). In most human cell studies, transgene integrations were typically reported to be single or low copy number events[45].
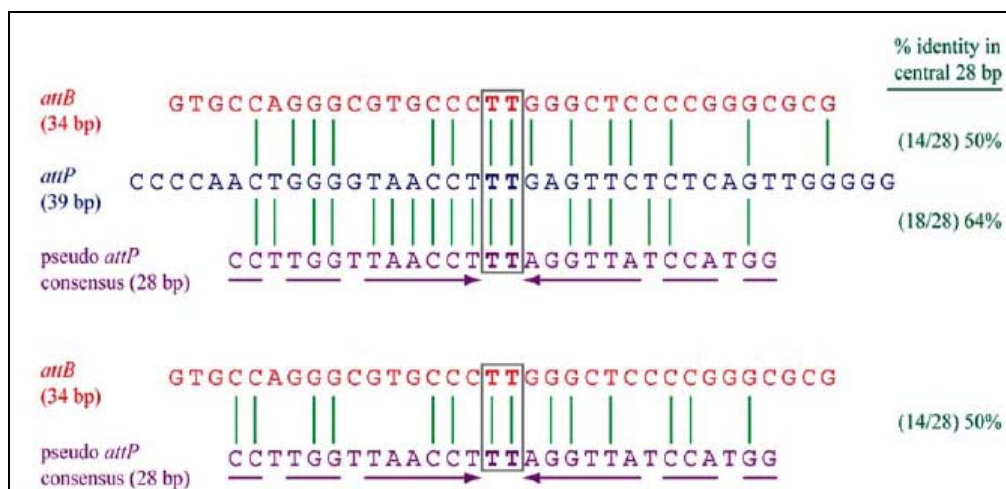
Figure 1.3.3   Sequence similarities between *attB*, *attP* and pseudo *attP* sequences. The minimal length wild type *attB* sequence (34 bp; depicted in red), wild type *attP* sequence (39 bp; depicted in blue) and pseudo *attP* consensus sequence (28 bp; depicted in purple) are aligned and compared to estimate the percent identity between them. The percent identity to the central 28 bp pseudo *attP* consensus sequence is given on the right. Wild type *attB* has 50% identity to both wild type *attP* and pseudo *attP* consensus sequences. Identity between the two latter sequences (wild type and pseudo *attP* consensus) is 64%. Green lines indicate sequence similarities. The central TT core sequence where integration or cross-over occurs is boxed. (*Figure taken from ref.44*).

The phiC31 integrase system has been effectively employed in RMCE studies to insert transgenes into pre-integrated wild type *attP* sites and also, more importantly, for stable gene transfer into endogenous pseudo *attP* sites in mammalian genomes. Its ability to mediate irreversible unidirectional site-specific recombination

into a limited number of chromosomal sites in human cells spurred intense interest as a relatively safer method for stable gene transfer for clinical applications. PhiC31 integrase has been successfully employed both *in vitro* and *in vivo* to induce stable expression of therapeutic transgenes. Ortiz-Urda *et al*. demonstrated functional correction of type VII collagen deficiency and laminin V deficiency in skin samples from patients with recessive dystrophic epidermolysis bullosa[46] and junctional epidermolysis bullosa[47], respectively. Thyagarajan *et al*. generated ES lines with stable transgene expression[48] and Ishikawa *et al*. showed the possibility of correcting X-linked SCID deficiency by expressing *IL2* receptor gamma chain in T cell-lines from SCID-X1 patients[49]. Successful correction of deficiencies of fumarylacetoacetate hydrolase[50], factor IX[51] and dystrophin[52] have also been demonstrated in murine models. Experimental data and bioinformatic analyses have suggested that phiC31 integrase could potentially mediate integrations into 370 different genomic sites, (202-764 sites based on 95% confidence interval) [45]. Furthermore, several studies have highlighted potential hot spots within mammalian genomes. For instance, integrations into 8p22 [53] and 19q13.31 [45] sites were frequently observed in human cells while frequent integrations into the murine locus, *mpsl1*, has also been reported often [53]. The limited number of potential sequence-specific integrations coupled with the possibility for long term transgene expression suggests that phiC31 integrase could be a safer alternative to randomly integrating vectors. However, several studies have raised the possibility that phiC31 integrase may induce infrequent chromosomal translocations[54, 55], possibly by promoting recombination between two endogenous pseudo *attP* sites in different chromosomes. Our data[56] suggest that the frequency of chromosomal aberrations is not fixed but rather may be influenced by the conditions in which the integrase acts e.g. may vary with different cell types. Using spectral karyotyping, we observed translocations in only 4 of 300 metaphases of primary cells treated with phiC31 integrase, a frequency similar to the low background of chromosomal abnormalities reported in normal human somatic cells[57]. Moreover, chromosomal translocations have been observed *in vitro* in cells treated with vectors already approved for clinical trials such as the rAAV vector[58], albeit without any pathological consequences *in vivo*. Concerns of potentially pathogenic chromosomal rearrangements have somewhat dampened interest in phiC31 integrase as an agent to be translated into clinical therapy. Although there is still a push develop gene therapy vectors with impeccable safety profiles, our work suggests that phiC31 integrase has a relatively benign biosafety profile compared to randomly integrating retroviral and lentiviral vectors. Attempts to increase the site-

specificity of phiC31 integrase include mutagenised versions which display increased bias for integrating in pseudo *attP* sites in chromosome 8p22[59] or other genomic sites[60], and versions with higher integration frequencies[61].

Thus, *ex vivo* gene therapy approaches utilizing phiC31 integrase could be rendered even safer by using integrases with greater site-specificity and pre-screening gene modified cells, preferably with high-throughput genome-wide methods, to exclude suspect cells and select cells with safe characteristics.

1.3.1.2. Targeted gene integration

Although transposases and SSRs integrate vectors non-randomly, some have questioned if these systems are truly sequence-specific or merely quasi-random as these systems are known to mediate integrations into degenerate sequences with very little homology to wild type sequences. The terms site-directed or targeted gene integration describe modifications designed to direct integration to specific genomic regions recognized by the modifying agent, usually a DNA-binding protein (DBP). Altering or skewing the integration preference of SSRs towards a particular locus is considered an advantage as it reduces the risk of off-target integrations into unsafe genomic regions. Gene targeting can be mediated by DNA-protein interactions or DNA-base pairing interactions. Naturally occurring DNA-binding proteins such as zinc finger proteins (ZFP) or viral peptides such as Rep have been deployed to favor DNA-protein interactions defined by their inherent specificities. Short oligodeoxynucleotides or short regions of homology have also been used to achieve homologous recombination in targeted regions (*vide infra*).

Several strategies have been proposed to achieve targeting specificity with DBPs. One approach is to tether a DBP to a recombinase by direct fusion or protein-protein interactions. This has the theoretical effect of enhancing local concentrations of the SSR at sites specified by the DBP and could more effectively restrict integration activity to a specific genomic region of choice. Care should be taken to ensure that the tethered SSR is not adversely compromised functionally. Another less frequently investigated approach relies on binding of a DBP to the vector sequence as a means of targeting vector sequences to the locus of interest[62]. The following sections review examples of targeted gene integration.

1.3.1.2.1. Targeting *via* DNA-binding proteins

A classical example of targeted gene integration are AAV vectors which have been reported to mediate 70 to 85% of integrations into the AAVS1 site in human chromosome 19q13.3. Site-specific integration of AAV requires viral Rep proteins

(Rep68/Rep78) that recognize Rep-binding elements (RBE) in the ITRs of AAV and in the AAVS1 genomic site[63]. This has led to the development of non-viral gene targeting using vector sequences flanked by AAV ITRs that can be recognized, nicked and integrated into AAVS1 sites by Rep proteins expressed in *trans* [64]. Philpott and collaborators[65] reported that a 138 bp P5 integration efficiency element within the ITR was sufficient for efficient Rep binding. More recently Feng *et al.*[66] showed that efficient RBE binding and targeted integration into AAVS1 could be achieved with vector sequences flanked by a 16 bp fragment within the ITR (RBEitr). Rep-based non-viral systems mediate AAVS1-specific integrations in *in vitro* clonal cultures[64, 67, 68] at frequencies ranging from 12 to 60%. On this basis, these systems have been tested and shown to function also *in vivo* [69]. In this sense, Rep protein may be regarded as a DBP that redirects vector sequences to a targeted genomic locus, notwithstanding the possibility of concurrent random integrations. The persistent potential for random gene integrations coupled with the need for antibiotic selection to isolate cells with the desired targeted integrations and the relatively low targeting efficiencies are possible reasons why this integrating strategy has not garnered much interest. Current versions of rAAV vectors do not express any viral proteins and hence are not intended to be integrating.

Several groups have explored the possibility of combining the integration mechanisms of transposons, HIV-1 integrase, phage integrase or SSRs with the desired DNA-binding specificities of DBPs. Early gene targeting studies relied on the use of a handful of well studied naturally occurring DBPs such as yeast Gal4 (binds upstream activating sequences), *Escherichia coli* Lex A (binds to Lex A operator sequence)[70], phage λ repressor (binds phage λ operator sites)[71] and murine transcription factors such as Zif268[72]. Although Gal4, lex A and λ repressor proteins were instrumental in demonstrating the feasibility of targeted gene integrations *in vitro*, they were not adaptable to clinical applications as they lack natural binding sites in the human genome. However, they have been used to bind vector sequences bearing their recognition elements and, fused with other endogenous DBPs, can be engineered to recognize elements in the human genome[20]. Other naturally occurring cellular DBPs, such as scaffold attachment factor (SAF)[20] and lens epithelium-derived growth factor (LEDGF)[21], also bind to several human genomic regions (albeit without precise sequence recognition) and facilitate integration *in vitro*. Recent work by Gijsbers and collaborators showed the potential for redirecting lentiviral integrations into transcriptionally inactive regions by modifying the natural LEDGF/p75-viral integrase interactions[73]. Such retargeting strategies could

potentially be adapted to engineer hybrid viral vectors with safer integration characteristics compared to current generations of viral vectors.

Amongst transcription factors, the ZFPs are an especially favored class of DBPs, given that the human genome codes for an estimated 4500 ZFPs[74]. An inherent limitation of naturally occurring ZFPs is their tendency to recognize short DNA sequences which may be present at many sites in the genome. This prompted engineering artificial ZFPs that could be tailored to bind to unique genomic sites. The E2C-ZFP was one of the first synthetic ZFPs that was designed to bind to a unique sequence in the 5' UTR of the *ERBB2* gene. Tan *et al*. demonstrated a 10-fold enrichment of integrations into the E2C binding site in human cells transduced with a HIV-1 integrase fused to E2C-ZFPs[75]. Advances in protein structure analysis and high-throughput techniques for testing DNA-protein interactions have ushered in new possibilities of creating user-defined custom ZFPs to target specific loci in the human genome. High expectations of the practical utility of customized ZFPs have spawned commercial investment in this technology, forming the business platform of Sangamo Biosciences which focuses on designing novel customized synthetic ZFPs as modulators of transcriptional control and as gene targeting agents in combination with nucleases (i.e. zinc finger nucleases). These artificial ZFPs could potentially redefine the integration spectrum of SSRs and viral integrases to enhance their biosafety.

Although tethering DBPs to recombinases and transposases has enriched targeted gene integrations, such chimeric systems continue to suffer from the disadvantage of non-directed integrations owing to residual activity of the recombinase/transposase and its inherent specificity. The holy grail of gene targeting is integration exclusively at a single user-defined safe harbor that does not incur the disruptive consequences of insertional mutagenesis/oncogenesis. This ideal may now be within reach with the advent of synthetic ZFPs, although the combination of such synthetic ZFPs with existing recombinases and transposase has not yet been rigorously evaluated. Recent years have also seen the development of other gene targeting systems based on homologous recombination which promise highly accurate gene integration but whose effectiveness has yet to be proven.

1.3.1.2.2.   Site-specific homologous recombination

The transgene integration strategies discussed thus far rely on the activity of an enzyme or protein to direct and mediate the integration of vector DNA into the genome randomly or with limited specificities. Another highly site-specific strategy that has been utilized for many years to create transgenic cells and animals with

targeted genome modifications exploits endogenous repair mechanisms of host cells to execute homologous recombination, thereby incorporating exogenous DNA into specific genomic sites. Effective homologous recombination requires transgenic DNA to be flanked by sequences homologous to the genomic sequences into which they are to be integrated. These exogenous DNA sequences are templates in the process of homologous recombination and are subsequently replicated along with the genomic locus during host cell divisions. The basal frequency of homologous recombination involving exogenous DNA is very low, occurring in 1 out of 10^5 - 10^7 treated cells. However, this frequency can be enhanced 1000-fold by creating site-specific nicks in the genome[76], thereby stimulating DNA repair at these sites. DNA is repaired by one of two main mechanisms i.e. non-homologous end joining (NHEJ) or homologous recombination (HR), although variations of these mechanisms are also possible. Error prone NHEJ results in genomic DNA repair without transgene integration while HR may result in site-specific integration of the transgene into the desired locus. In the context of gene therapy, the prospect of exploiting homologous recombination is appealing as it holds the potential for targeted gene repair and precise transgene integration into safe genomic loci. A patient's cells could in theory be modified *ex vivo* to correct disease-causing mutations or to integrate a transgene for long term expression of a deficient or defective protein before reimplanting into the same patient (autologous cell therapy). Recent advances exploring such strategies will be discussed in this section.

1.3.1.2.2.1.   Meganucleases

A more efficient and reproducible strategy for gene editing or integration that has been the focus of recent research is the use of highly site-specific endonucleases to induce double-stranded DNA breaks in specific genomic sites to stimulate deletions *via* non-homologous end joining or homologous recombination of exogenously delivered DNA into these sites. Three main classes of engineered endonucleases have emerged: zinc finger nucleases which are chimeras of ZFPs and the catalytic domain of *Fok* I restriction endonuclease; chemical endonucleases which consist of chemical or peptidic cleavers fused with DNA recognizing polymers; and meganucleases (homing endonucleases) which are capable of recognizing and cleaving target DNA sequences, usually 14 -40 bp in length. HO endonuclease which mediates mating type switch in *Saccharomyces cerevisiae*, I-CreI and I-SceI meganucleases are examples of naturally occurring homing endonucleases. However, applications of naturally occurring meganucleases have been limited either by the lack of recognition sites or by the presence of more than a single site in the human

genome. The LAGLIDADG family of meganucleases includes I-SceI and I-CreI which are the largest and best characterized meganucleases, and are active as monomers or homodimers. Their catalytic cleavage centers are embedded within the DNA-binding domains, thus making non-specific cleavage very unlikely. Elucidation of the protein structures of endonucleases such as SceI and CreI have accelerated engineering of meganuclease variants with unique genomic recognition sites. Most effort has been directed at developing I-CreI and I-SceI variants with unique specificities and reduced off-target cleavage activity. Thus far two engineered meganucleases cleaving unique genomic loci in the human *XPC* [77] and *Rag1* genes [78] have been reported. Other improvements have been to engineer variant CreI (naturally homodimeric) meganuclease to function as obligate heterodimers[79] or as single-chain derivatives[80]. Computational approaches[81] have integrated structural and high-throughput screening data to identify the cleavage properties of 18000 engineered meganucleases, based mostly on CreI meganuclease[82].

Thus far, homologous recombination involving transgenes with meganucleases has been demonstrated in only a few cell types and a comprehensive evaluation of their genotoxic potential is awaited. The future development of engineered variants that collectively offer a wide spectrum of unique integration sites may be useful but will need careful evaluation. At present, there is a need to engineer endonucleases for user-defined specificities. This requirement may be more readily fulfilled with zinc finger nuclease technology given the potentially broader spectrum of genome-specific ZFPs that can be custom engineered.

1.3.1.2.2.2.   Zinc finger nucleases

Zinc finger nucleases (ZFNs), first conceptualized by a collaborative effort between the groups of Chandrasekaran and Carrol[83], are synthetic chimeras composed of a tandem array of DNA-binding Cys2-His2 zinc finger proteins fused with the catalytic domain of *Fok*1restriction endonuclease *via* a short linker peptide. The Cys2-His2 zinc finger peptides, typically made up of 30 amino acids and usually containing two cysteine and histidine residues coordinated by a single zinc atom to form a βßα structure, are capable of specifically recognizing and binding to 3 consecutive bases of DNA. Amino acids in the N-terminus of the α-helix interact specifically with bases in the major groove of DNA, with amino acids in positions -1, 3, and 6 of each zinc finger peptide [84] (numbered relative to the start of helix) making contact with the 3', middle and 5' nucleotide bases in the recognized triplet sequences. In addition to these interactions, there is contact between an aspartate

residue in position 2 of the ZFP with either an adenosine or cytosine on the complementary strand immediately preceding the triplet recognition sequence of the ZFP[84] (**Figure 1.3.4**). ZFNs are typically designed as a pair, each consisting of 3 to 6 ZF monomers recognizing 12 to 18 consecutive bases on each strand with a spacer region of about 5 to 7 bases between the neighboring ZFN pair targeting sequences on complementary DNA strands. The full DNA recognition site for a given pair of ZFNs is given by sequences encompassing the two ZFN half sites and spacer region, and is typically around 30 to 40 bases in length. Theoretically, a 30 bp sequence occurs only once per 1.15×10^{18} nucleotides. Given that the human genome is made up of approximately 3×10^9 bases, the occurrence of a 30 bp sequence ought to be unique. *FokI* endonuclease is only active as a dimer, being non-functional in its monomeric state, as is the case where only a single ZFN binds to its target site. The requirement for dimerization of two monomeric *Fok*I nucleases from each protein of a ZFN pair ought to ensure DNA cleavage only when both proteins of a ZFN pair bind to the correct target sequence. The ensuing dimerization and activation of *FokI* nuclease then induces a double-stranded DNA break within the spacer region between the two ZFN proteins. First generation ZFNs formed *Fok*I homodimers that bound to similar adjacent half sites, thus resulting in off-target *FokI*-induced DNA cleavage that caused significant cellular toxicity. This has been partly overcome by redesigning the *Fok*I dimerization domains to form obligate heterodimers. Work by two separate groups showed that different amino acid substitutions within the *Fok*I domain of each protein in a ZFN pair generated electrostatic repulsion of between identical *FokI* domains, thereby minimizing ZFN homodimers and reducing their off-target effects [85, 86].
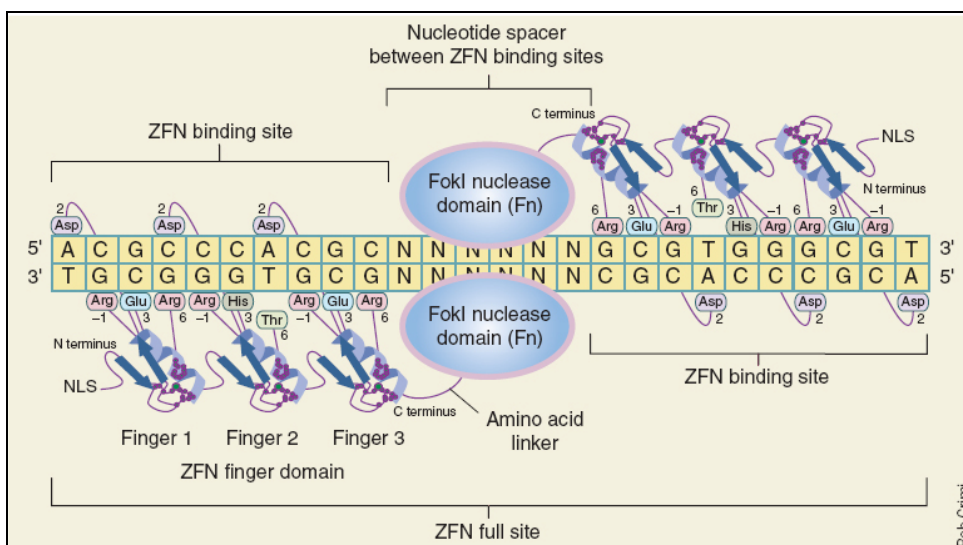
Figure 1.3.4   Schematic of ZFN binding to target sequence. ZFNs consist of a tandem array of Cys2His2 zinc finger peptides fused at the C-terminal *via* a short linker sequence to the catalytic domain of monomeric *Fok*I endonuclease. ZFNs are designed as protein pairs that bind to adjacent DNA sequences on complementary DNA strands and are usually separated by 5 to 7 bases in order to allow for dimerization of *Fok*I monomers. The dimer creates a double-stranded DNA break within this spacer region. DNA binding specificity is determined by the amino acid sequences of each ZFP that specifically recognizes and binds to 3 consecutive bases. Amino acids in position -1, 3 and 6 within the α-helix of each ZFP with a ββα structure typically make contact with the 3', middle and 5' bases, respectively, of the recognized triplet within a single DNA strand. There is an additional interaction between an aspartate residue in position 2 of the ZFP with either an adenosine or cytosine base in the complementary strand immediately preceding the triplet recognition sequence. The full ZFN recognition site is given as the sequence encompassing two ZFN binding sites plus the spacer sequence. (*Figure taken from ref.84*).

The design of site-specific ZFNs is based on naturally occurring zinc finger transcription factors such as the murine Zif268 or human specificity protein 1 (SP1) which provide the scaffold in which each Cys2-His2 zinc finger that specifically recognizes a base triplet can be replaced to derive a novel ZFP capable of binding to unique genomic sequences of choice. Such polydactyl ZFP have been assembled by modular assembly [87] in which individual ZFs are combined as a succession of modules to form an array capable of recognizing a length of DNA sequence. Alternative strategies such as oligomerized pool engineering (OPEN) [88] or context-dependent assembly (CoDA)[89] take into account the context dependence of each individual ZFs relative to its neighboring partners in sequence recognition and binding. The main function of ZFNs is to function as highly site-specific molecular scissors to create double-stranded DNA breaks at user-specified genomic loci. Double-stranded DNA breaks (DSDB) in the genome are systematically detected by intracellular DNA damage sensing proteins and ultimately repaired by NHEJ or homology directed repair (HDR). These repair mechanisms are crucial for maintaining the genomic integrity of cells that constantly encounter exogenous or endogenous stresses that damage DNA.

NHEJ is one of the key processes that repairs double-stranded DNA breaks in the genome. NHEJ is thought to be the major pathway of repair and is most active in the G_0-G_1 and early S-phase of the cell cycle[90]. The NHEJ pathway of repair can simply be summarized as a homology independent process which joins two ends of a double-stranded break, often creating small deletions or insertions at the sites of repair. The NHEJ pathway has been well elucidated and involves key DNA damage sensing and repair proteins (**Figure 1.3.5**). Very briefly, the Ku70/Ku80 (Ku) proteins are responsible for binding to DNA termini and aligning the DNA ends for ligation.

DNA-PK$_{cs}$ (DNA-dependent protein kinase catalytic subunit) recruited to DNA-bound Ku heterodimers at damaged ends, together with Artemis protein, stimulates processing of DNA ends. Recessed DNA ends are finally joined together by the *XRCC4* (X-ray repair complementing defective repair in Chinese hamster cells 4)-DNA ligase IV complex to complete the NHEJ repair process[91].
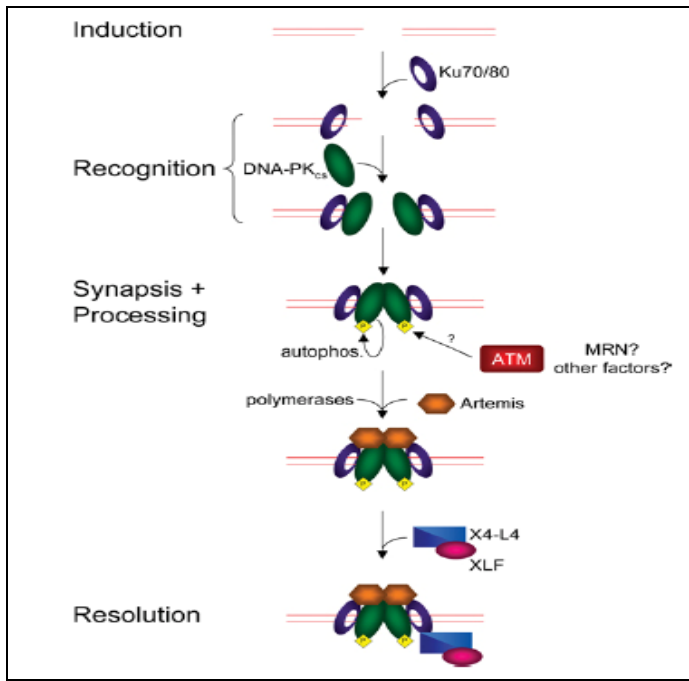
Figure 1.3.5   NHEJ pathway. Schematic summarizing the key proteins involved in the NHEJ pathway of DNA repair. Induction of repair of resected ends of DNA is initiated by binding of Ku70/80 heterodimers which in turn recruit *DNA-PKcs*. Upon binding, *DNA-PKcs* is activated by autophosphorylation. This allows Artemis and other proteins to bind to the DNA repair complex and stimulate processing of DNA ends. X4-L4 complex consisting of *XRCC4* and DNA ligase IV together with *XLF* mediate ligation of the processed DNA ends to repair the DSDBs. (*Figure taken from ref.91.*)

HR or HDR is another key DNA damage repair process during which information from sister chromatids or exogenously provided homologous DNA templates are copied into the newly synthesized DNA strand, ensuring that integrity of DNA sequences in the repaired strands is maintained[92] (**Figure 1.3.6**). Unlike the NHEJ pathway, HR or HDR is usually results in error-free and precise repair, and functions more prominently in the late S- to G2-M phases of the cell cycle [93]. The MRX (*MRE11-RAD50-XRS2*) complex is thought to play an important role in HDR pathway. In HDR repair, the MRX complex binds to break sites and together with a 5'-3' resection exonuclease converts the double-stranded (ds) DNA ends to single-stranded (ss) DNA with 3'-hydroxyl overhangs. Secondary structures on resected ssDNA are eliminated by binding of RPA (replication protein A) to the ssDNA.

Rad52 binds to the RPA-coated ssDNA and recruits Rad51 which extends onto ssDNA mediated by Rad55/Rad57 and, in the process, displaces RPA from the ssDNA. The Rad51 nucleoprotein complex is also thought to be involved in homology search for the repair DNA template. *Rad54*, *via* its interaction with *Rad51*, promotes chromatin remodeling, DNA unwinding and strand annealing between the *Rad51* nucleoprotein coated ssDNA and a suitable homologous DNA repair template. DNA synthesis using homologous DNA as the repair template ensues, followed by strand displacement or resolution. Although there are several different models to explain the resolution of the newly synthesized DNA, it is believed to occur mainly through the synthesis dependent strand annealing (SDSA) pathway in somatic mammalian cells. In SDSA, the newly synthesized strand displaces from the repair template and anneals with the other end of the DNA break [91]. HR or HDR thus serve as a mechanism to correctly repair a DNA break in the genome and can be exploited to copy exogenous DNA sequences (flanked by homologous sequences) into a particular genomic locus for site-specific transgene integration.
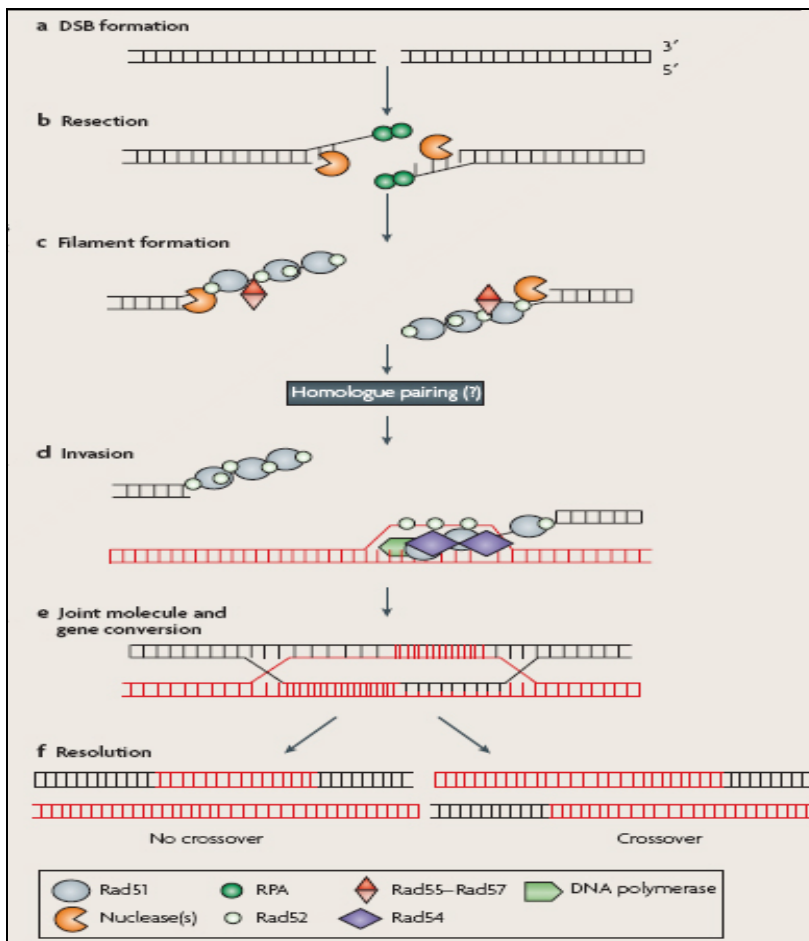
Figure 1.3.6 HDR pathway. DSDBs are recognized and bound by the *Mre11-Rad50-Nbs1* (MRN) complex. Human exonuclease I (*hExoI*) resects the broken ends of DSDBs to convert them to 3' single-stranded ends which are then bound and

stabilized by ssDNA binding protein, replication protein A (*RPA*). By a process mediated by *Rad52*, *Rad55* and *Rad57*, *Rad51* replaces *RPA* on the resected single-stranded DNA, thereby forming a nucleofilament that participates in homology search and homologue pairing with sequences in either a sister chromatid or an exogenous DNA template. *Rad54*, belonging to SWI-SNF family of helicases, is thought to induce an open chromatin configuration and participate in strand invasion and dissociation of *Rad51* from bound DNA. The invading DNA strand results in a D-loop structure that primes DNA synthesis by DNA polymerases. Newly copied DNA sequences may subsequently be resolved with or without crossover of genetic material between homologous DNA strands. (*Figure taken from ref.92.*)

Two major therapeutic applications of ZFNs are permanent gene disruption by creating insertions/deletions during error-prone NHEJ repair or site-specific transgene insertion *via* repair by HR or HDR (**Figure 1.3.7**).
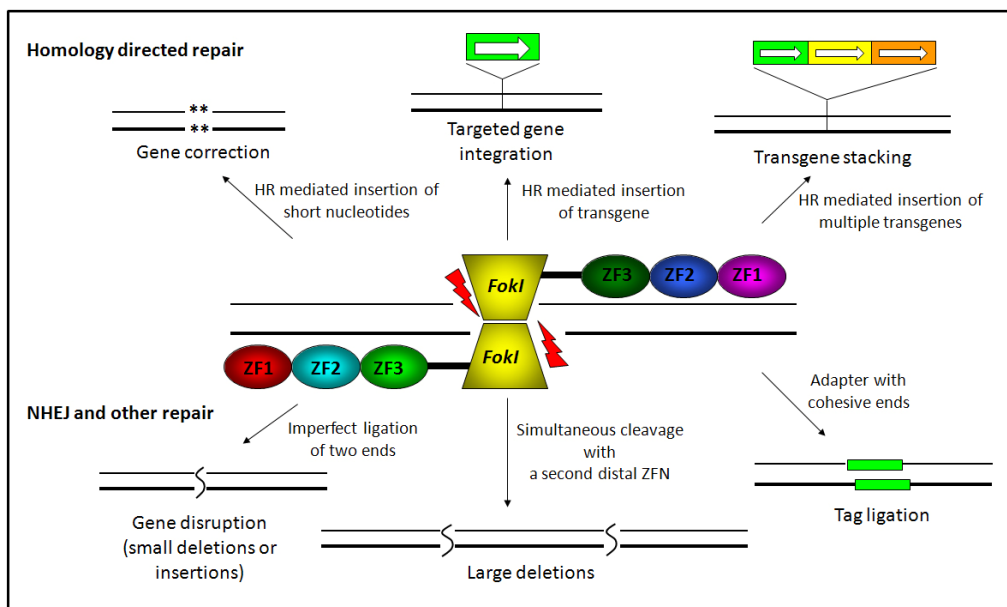
Figure 1.3.7   Genome editing with ZFNs. Site-specific cleavage of genomic DNA by ZFNs can be repaired by homology-directed repair to correct or induce point mutations, or to insert single or multiple transgenes (in the presence of donor DNA). Repair by NHEJ results in gene disruption caused by small insertions and/or deletions. Site-specific insertion of molecular tags and generation of large genomic deletions may also be achieved with ZFN-mediated cleavage of genomic DNA. (*Adapted from Urnov FD, et al., 2010.*)[94]

Since the turn of the millennium, ZFN technology has been harnessed to demonstrate feasibility of targeted gene corrections, transgene insertions and gene disruptions, in addition to pioneering a new approach for deriving transgenic plants and animals. ZFN technology has been used to derive transgenic crops with improved traits by mutagenesis of genes or targeted integration of herbicide resistance genes in species such as *Arabidopsis thaliana*, *Nicotiana tabacum and Zea mays*[95, 96] and to derive specific gene knock-out strains of mice and rats[97]. Given the ability to permanently disrupt specific genes, ZFNs have proved useful for elucidating gene

functions during embryogenesis and development. Heritable targeted gene disruption has been demonstrated in human embryonic stem cells, *Danio rerio* and *Drosophilia* [98-100]. ZFN-mediated gene knock-out has been effectively employed to disrupt the C-C chemokine receptor type 5 (*CCR5)* locus in human HSCs as a possible therapeutic strategy to confer resistance to HIV-1 infection by adoptive cell therapy *in vivo* [101]. The use of ZFN-modified T-cells is currently being tested in three phase I human clinical trials - for HIV-1 treatment (NCT00842634, NCT01044654) and glioblastoma (NCT01082926)[94]. Targeted disruption of several other genes such as *Bax* and *Bak* has also been shown in human cells[102]. More recently, Liu *et al*. showed the feasibility of generating triple gene knock-outs in cell-lines using ZFNs[103]. The ability to correct single-base genetic mutations and the theoretical potential for exquisitely precise site-specific gene insertions has opened a plethora of possibilities for gene therapy applications. Porteus and Baltimore first reported the possibilities of targeted ZFN-mediated genome editing in human somatic cells with gene correction of a pre-integrated GFP reporter gene[104]. Work by Urnov *et al*. has also been influential in demonstrating efficient correction of a *IL2Rγ* gene mutation in human cells, pointing to the prospect of future therapy for SCID-X1[105]. Others have shown the feasibility of integrating exogenous DNA up to 8 kb into the same locus [106], and other human genomic genes such as *PIGA*, *PPP1R12C* and *POU5F1* in primary cells such as mesenchymal stem cells (MSCs)[107], cord blood derived CD34+ HSCs[108], embryonic stem cells and induced pluripotent stem cells (iPSCs)[98, 109].

A current limitation of ZFN technology for site-directed transgene insertion is concern about unintended genomic modifications and possible biological hazards therefrom. Although several groups have demonstrated that the likelihood of off-target genomic modifications is low, there has been no comprehensive genome-wide analysis to date to rigorously support these claims. Potential off-target interactions of ZFNs must be evaluated by genome-wide techniques such as array-based methods combined with deep sequencing in order to detect rare integration events. Long term monitoring of ZFN-modified cells is essential, using small and large animal models to assess fully any potential genotoxicity. The current efficiency of targeted gene insertion using ZFNs is still relatively low and may not warrant its broad application in human gene therapy. This awaits more specific ZFNs with robust and efficient genome targeting activity. Several useful resources are currently available in the public domain to aid the design, construction and testing of specific ZFNs. Helpful information and software tools pertaining to ZFN design and construction as well as a collection of ZFN plasmids and reagents for constructing and testing ZFNs are readily

available to the research community at The Zinc Finger Consortium (http://www.zincfingers.org). Information on individual C2H2 zinc fingers and engineered zinc finger arrays have been compiled into databases such as the Zinc Finger Database[110] (ZiFDB; http://bindr.gdcb.iastate.edu/ZiFDB). Web-based resources such as Zinc Finger Targeter[111] (ZiFiT; http://bindr.gdcb.iastate.edu/ZiFiT/) and more recently ZFNGenome[112] (http://bindr.gdcb.iastate.edu/ZFNGenome) and ZFN-Site[113] (ccg.vital-it.ch/tagger/targetsearch.html) provide excellent tools to aid the identification of potential ZF binding sites in user supplied target regions. They include software that calculates strengths of predicted ZFNs to be engineered by modular assembly or the OPEN method, and also give information regarding potential off-target binding sites. Furthermore, Sangamo Biosciences and several other groups have described assays to evaluate the functional specificities of user-designed ZFNs. Recent improvements to ZFNs have also used *Fok*I variants with increased cleavage activities, in an attempt to increase the rate of genome modifications[114, 115]. Higher ZFN cleavage activity, possibly due to increased protein stability, was also achieved by conditioning cells to transient mild hypothermia[116]. Recently, ZFN variants (ZFNickases) that induce single-stranded nicks instead of double-stranded DNA breaks have been engineered and shown to suppress DNA break repair by the error-prone NHEJ pathway, thus increasing the frequency of HR mediated repair [117, 118]. Although ZFNickases seem, their significantly lower efficiency of gene integration compared to ZFNs limits enthusiasm at the current stage of their development. Further improvements to ZFNickases are necessary before they can be considered serious candidates for targeted gene integration applications.

We need to better understand the factors that influence the efficiency of homologous recombination and learn how to exploit them to increase gene targeting efficiencies to levels that are clinically meaningful. More work is needed to identify and test safe harbors in the human genome and to design ZFNs targeting them. Lastly, improvements to vector designs such as CpG-free vectors, the use of suitable physiological promoters, codon-optimized transgenes and incorporation of relevant insulator and enhancer elements would be pertinent features to promote durable transgene expression and minimize risks of insertional gene mishaps.

An ideal gene-based treatment for some monogenic disorders would be to derive self-renewing cells expressing a corrected version of the defective gene *via* site-specific integration in a safe genomic locus. Such gene modified cells could be exhaustively evaluated for their genotoxic potential *ex vivo* before being administered

into patients. Given the lexicon of site-specific ZFNs that is being developed, this could be a real possibility in the near future with ZFN-modified stem cells.

1.3.1.2.2.3.   TALENs

Transcription activator-like effector nucleases (TALENs) are a recently introduced class of site-specific nucleases that function in plant and mammalian cells[119]. Transcription activator-like effector (TALE) proteins were first identified as pathogenic proteins, secreted by plant pathogens of the *Xanthomonas* genus, that were able to modulate host gene expression in infected plant cells by acting as DNA-binding transcription factors. TALE proteins are made up of highly conserved and modularly assembled 33- to 35- amino-acid repeat sequences, known as TALE repeats or monomers, capable of binding DNA specifically. Naturally occurring TALE proteins are composed of 1.5 to 33.5 repeat elements with an average of 34 amino acids per TALE repeat [120]. Two variable amino acids in the 12th and 13th positions of each TALE repeat, known as repeat-variable di-residues (RVD), bind specifically to a single DNA base and confer the DNA-binding specificity to each TALE monomer. In 2009, back to back publications in Science reported the successful deciphering of the "TALE codes" that govern the DNA-binding specificity of TALE proteins. Two groups demonstrated that DNA-binding specificity could be modulated by changing the RVD sequences within each repeat element of modular TALE proteins [120, 121]. Following the successful precedent of combining zinc finger peptides with nucleases for targeted genome modifications, attempts were made to develop similar site-specific nucleases composed of TALE proteins. TALENs were first successfully engineered by combining modularly assembled TALE repeat elements with the catalytic domain of *FokI* nuclease [122]. TALENs are typically designed as pairs to recognize and bind to between 12 to 24 bases of DNA on opposing DNA strands (with a separating spacer of around 14-20 bases). Like ZFNs, TALENs allow dimerization and activation of *FokI* nuclease monomers, enabling generation of DNA double-stranded breaks at user-specified sites in the genome. Given their highly repetitive modular structure and considerably larger size compared to ZFNs, cloning TALENs proved to be more challenging. Some of these technical limitations have now been overcome by the Golden Gate cloning method[123] and fast ligation-based automatable solid-phase high throughput (FLASH) systems [124]. The number of genomic loci and genes successfully targeted and modified with site-specific TALENs has steadily increased over the few years since the inception of TALENs. The major advantage of TALENs is the ability to design TALENs targeting

practically any possible DNA sequence in the genome compared to ZFNs which are restricted by rules governing successful design of efficient ZFNs, such as the requirement for 5'-GNN-3' and 5'-ANN-3' triplet sequences [125]. Some preliminary results also suggest that TALENs are as efficient as ZFNs in site-specific cleavage, while having less off-target activities[126] . In summary, although TALENs are a more recent innovation than ZFNs and have not been extensively evaluated, they appear to have considerable potential as genome editing agents for gene- and cell-based therapy.
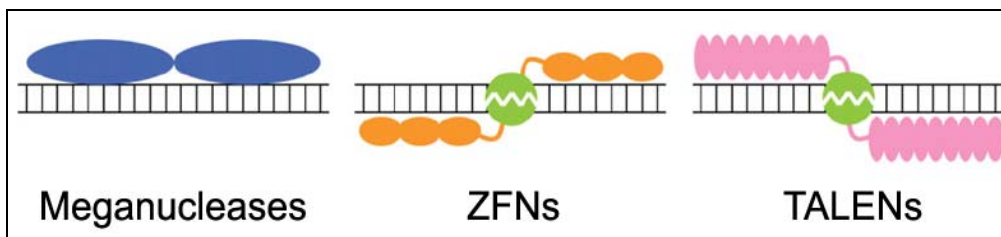
Figure 1.3.8 Schematic of site-specific nucleases Meganucleases (blue) bind to target sites (12-18 bases) as dimers. Heterodimeric ZFNs, composed of tandem arrays of ZF peptides (orange) fused with monomeric *FokI* nuclease (green), bind to adjacent target sites (separated by 5 to 7 bases) on opposing DNA strands (each ZF peptide binds 3 consecutive bases). TALENs, composed of tandem arrays of TALE repeats (pink) fused with monomeric *FokI* nuclease (green), bind to adjacent target sites (separated by 12-30 bases) on opposing DNA strands (each TALE repeat binds to a single base). (*Figure taken from ref.127.*)[127] .

1.3.2. Potential safe harbors in the human genome for transgene integration

While stable expression of transgenes from genomic integration appears to be an appealing objective for gene therapy, iatrogenic oncogenic complications due to insertional mutagenesis reiterate the real dangers associated with random integration and emphasize the need to direct transgene integrations into safe regions within the genome. Although there are no standard definitions of a "safe genomic harbor", given the untoward consequences of genomic disruptions, one can attempt to propose characteristics of a safe harbor. Thus, it should be a locus that is distant or insulated from endogenous genes whose altered expression is known to be associated with diseases, or to induce abnormal cell physiology and/or genome instability. Although it would be appealing to envisage non-coding, non-gene rich or extragenic regions as good candidates for integrations, these regions are often associated with heterochromatic signatures which may not be permissive for transgene expression. Heterochromatin may not be readily accessible to transgene integration processes, and even in the event of integration, may silence transgene expression. The role of non-coding regions in gene regulation and their contribution to disease phenotype is

another reason to be circumspect in assuming non-coding regions to be general safe harbors. An important exclusion criterion of safe harbors is close proximity to oncogenes, tumor suppressor genes and other key disease-associated genes, whose perturbation as a consequence of integration could have highly regrettable consequences. Targeting an endogenous gene, if absolutely necessary, should be one whose disruption is known to have inconsequential effects. In summary a favorable genomic region for integration should ideally not disrupt the cellular transcriptome, should not perturb neighboring gene expressions and should allow for durable transgene expression without disrupting cellular functions or inducing adverse outcomes[128].

1.3.2.1. *CCR5* locus

The chemokine (C-C motif) receptor 5 gene (*CCR5*), encoding a co-receptor for HIV that is also a surface receptor for chemokines expressed by immune cells such as T cells, monocytes, macrophages and microglia, has been selected as a genomic site for targeted transgene integration. Initial interest in this locus on 3p21.31 arose because targeted disruption of *CCR5* could protect against HIV infection. As an extension of this idea and given that the *CCR5* gene is also expressed in non-immune cells, there have been attempts to integrate therapeutic transgenes into the *CCR5* locus for sustained expression. For instance, Benabdallah and co-authors, reported durable erythropoietin secretion from MSCs integrated with the erythropoietin (*EPO*) transgene at the *CCR5* locus [107]. However, it should be noted that the function of *CCR5* in non-immune cells such as neurons, endothelium and smooth muscle cells, and the effects of its disruption have not been well studied. A phase I clinical trial of autologous implantation of *CCR5* disrupted T cells into patients has thus far not reported any adverse consequences, suggesting potentially benign effects, if any, of disrupting endogenous *CCR5* gene expression.

1.3.2.2. Human ribosomal DNA

The human 45S ribosomal DNA (rDNA) which codes for over 400 copies of the 45S pre-RNA (rRNA) is clustered within the short arms of acrocentric chromosomes 13, 14, 15, 21 and 22. In humans, loss or gain of short acrocentric chromosomal arms has not been associated with phenotypic abnormalities, suggesting that targeting transgenes into these loci could be relatively benign. Furthermore, a considerable level of redundancy and capacity for compensation of rDNA gene function is likely given the presence of more than 400 copies of rDNA genes in the genome. Targeting these multiple loci could theoretically allow for multiple

integrations which would translate to higher overall transgene expression. Durable transgene expression would also be anticipated from integrations into rDNA sites given their transcriptionally active status. A handful of investigators have successfully demonstrated expression of transgenes, such as clotting factors VIII and IX (FVIII and FIX, respectively) in human cells by targeted integration into rDNA loci [129]. The main caveat is that there is actually no site-specificity and transgenes could be integrated in any of the 400 regions encoded by rDNA. Risks of an adverse event occurring could be related to number of integration events occurring within a cell. At present, gene targeting to rDNA locus still relies on using conventional homologous recombination targeting strategies, making it relatively inefficient.

1.3.2.3. Human *ROSA26*, *ENVY* and *HPRT* locus

The human *ROSA26* locus in chromosome 3p25.3, first identified by Irion *et al*., is the human homolog of the murine *ROSA26* locus which is a common locus for targeted transgene integrations, especially in studies of murine ES cells. Whilst targeted transgene integration and expression has been demonstrated in the human *ROSA26* locus [130], relatively little is known of the endogenous gene functions and consequences of its dysregulation. As the *ROSA26* locus is gene dense, it is essential to evaluate effects of transgene integrations on the expression of neighboring genes. More studies are needed to evaluate the feasibility of utilizing this locus for targeted transgenesis.

The *ENVY* [131] and *HPRT* [132] loci have been targeted in human ES cells for transgene integration and studies have shown durable transgene expression from this locus as well as normal differentiation of transgenic ES cells. Gene targeting events at the X chromosome-linked *HPRT* gene result in disruption of the hypoxanthine guanine phosphoribosyl transferase (*HGPRT*) enzyme. Such cells can be selected based on their resistance to 6-thioguanine and 8-azaguanine, followed by rescue of the inactivated purine salvage pathway with hypoxanthine-aminopterin-thymidine (*HAT*) selection media. While the functional consequences of *ENVY* gene inactivation is uncertain, disruption of the *HPRT* gene, at least in neuronal cells, is linked with the neurogenetic Lesch-Nyhan syndrome. While *ENVY* and *HPRT* loci may be suitable for ES cell studies, they may not fulfill the criteria of safe genomic harbors.

1.3.2.4. AAVS1 locus

The AAVS1 locus in chromosome 19q13.42 is another putative safe harbor in the human genome. The pseudonym AAVS1 originated from the observation that wild type AAV2 virus, in the presence of viral Rep68/78 proteins, frequently

integrates its provirus into a Rep78 binding site in chromosome 19. This region, later called the *AAVS1* site, is actually located within the promoter region of a poorly characterized gene, protein phosphatase 1 regulatory subunit (*PPP1R12C*) also known as myosin-binding subunit 85 (*MBS85*), whose function is thought to involve the regulation of actin-myosin fiber assembly[133]. The fact that a relatively large population of humans may have been naturally infected with AAV and possibly carry viral integrations at the AAVS1 site without pathological consequences has been used to support the notion that this site could indeed be a safe harbor for targeting transgene integrations[134]. The *PPP1R12C* gene is ubiquitously expressed in several cell types, suggesting an open chromatin configuration of the AAVS1 locus that could theoretically be amenable for gene targeting by allowing access to specific targeting proteins that interact with and modify this locus. Furthermore, a transcriptionally active locus should theoretically better support durable transgene expression in contrast to a transcriptionally inactive region where heterochromatinization would likely lead to transgene silencing. *In vitro* studies in cell lines, ES cells and primary cells have demonstrated durable expression of transgenes integrated into the AAVS1 locus[135]. The presence of native insulator elements[136] is also thought to contribute to sustained expression of integrated transgenes while also preventing trans-activation of neighboring genes. Studies performed thus far in ES[98] and iPS[137] cells suggest no gross abnormalities or differentiation deficits arising from transgene integration into the AAVS1 locus. Mono- and biallelic modifications of the locus have no reported functional consequences in the modified cells, although biallelic modifications disrupted endogenous *PPP1R12C* expression [138]. Thus far, the cellular and organismal consequences of disrupting endogenous *PPP1R12C* expression are unclear and not well elucidated. Given the foregoing caveat, the ability to support sustained transgene expression without known adverse consequences thus far makes the AAVS1 locus an appealing locus for gene targeting.

1.3.3. Suicide genes as safety mechanisms for treatment modalities

The benefits of gene therapy for life threatening diseases for which there is currently no effective or affordable treatment justify their continued evaluation in clinical trials despite the known risks of iatrogenic complications. It is clear from the preceding sections that most research efforts have been directed at enhancing the biosafety of gene therapy vectors. An additional strategy to intervene and reverse adverse vector effects is to incorporate secondary safety mechanisms capable of

rapidly triggering the selective elimination of rogue transgenic cells. Suicide gene therapy or gene-directed enzyme prodrug therapy relies on the expression of transgene products from "suicide genes" that convert inactive prodrugs into cytotoxic drugs, thus selectively eliminating transgenic cells that express the suicide gene. Several suicide genes such as *Herpes simplex* virus thymidine kinase (HSV-TK), bacterial cytosine deaminase (CD), bacterial carboxypeptidase-G2 (CPDG2), purine nucleoside phosphorylase (PNP) and nitroreductase (NR) and their cognate prodrugs have been tested for their efficacy as agents of selective cell destruction[139]. Problems such as suicide gene silencing, incomplete elimination of targeted cells, cytotoxicity to non-gene expressing cells and immune response to suicide genes have reduced the efficacy of such approaches. Continued improvements to existing suicide genes and prodrugs as well as development of novel genes capable of selective elimination of cells with reduced cytotoxicity to normal cells are necessary improvements to suicide gene therapy for clinical applications. Recent developments in suicide gene therapy strategies will be briefly discussed in this section.

1.3.3.1. HSV thymidine kinase

The HSV-TK suicide gene and its prodrug, gancyclovir (GCV) is one of the most extensively studied and the only clinically validated suicide gene/prodrug system. HSV-TK phosphorylates the non-toxic acyclic analogs of deoxyguanosine such as GCV and acyclovir (ACV) into a toxic form that becomes incorporated into DNA. This leads to eventual cell death by inhibiting DNA synthesis and disrupting DNA replication in sensitive cells. The use of HSV-TK has found broad applications *in vitro* as negative selection in homologous recombination studies and has been successfully used in phase I-II clinical trials for prevention of graft *versus* host disease following allogeneic stem cell transplantation[140]. It has also been investigated extensively in cancer gene therapy to eliminate tumor cells. An ongoing phase III clinical trial by Ark Therapeutics (www.arktherapeutics.com) is evaluating HSV-TK combined with surgery and chemotherapy in patients with high grade gliomas (cited by Preuß E, *et al*., 2010). However, there are certain disadvantages of the HSV-TK/GCV system. These include GCV toxicity at clinical doses, insensitivity of HSV-TK expressing cells to GCV due to inactive spliced HSV-TK variants[141], cellular toxicity of high levels of HSV-TK that phosphorylate endogenous thymidine[142] and the inherent immunogenicity of viral epitopes presented by HSV-TK protein[143]. Several improvements have been made to improve the performance of HSV-TK such as splice-corrected variants[144], improved GCV sensitivity[145] and decreased affinity for

endogenous thymidine[142]. Notable HSV-TK variants with improved sensitivity to GCV include the SR39[145] and Q7530A[146] mutants. Splice corrected versions of HSV-TK (scHSV-TK) have been derived by mutating internal splice sites within wild type HSV-TK gene to prevent the emergence of GCV-resistant cells expressing inactive HSV-TK splice variants[144]. Another recent development is the use of a codon-optimized HSV-TK A168H mutant, TK007, which causes faster and more robust GCV mediated killing of cells while having less non-specific cytotoxicity[147] due to its reduced affinity for endogenous thymidine. These improved versions of HSV-TK could function effectively as benign suicide genes that could be activated to selectively eliminate implanted gene modified cells in the event of a serious adverse complication e.g. oncogenic transformation. However, outstanding issues such as immunogenicity of HSV-TK and the possibility of immune-mediated rejection of gene modified cells reiterate the need to investigate other novel human-based and possibly non-immunogenic suicide genes as better alternatives.

1.3.3.2. Cytosine deaminase and thymidylate kinase

Another widely used system for gene-directed enzyme prodrug therapy is the bacterial enzyme, cytosine deaminase. Cytosine deaminase preferentially deaminates 5-fluorocytosine to 5-fluorouracil which is converted by cellular enzymes to fluorodeoxyuridylate, an irreversible inhibitor of thymidylate synthase. This leads to a block in dTTP synthesis, DNA replication arrest and ultimately to apoptosis. However, this system may be better suited for cancer gene therapy, such as eradicating cancer cells from tumor beds, rather than for selective elimination of gene modified cells given that the prodrug product is freely diffusible across cell membranes and displays a localized general toxicity or bystander effect on neighboring cells[148].

A novel suicide enzyme is human thymidylate kinase which phosphorylates and converts the prodrug 3'-azido-3'-deoxythymidine (AZT) into toxic AZT-triphosphate (AZT-TP) and inhibits DNA replication in eukaryotic cells. Engineered mutants of thymidylate kinase have over 200-fold increased activity of conversion and inducing apoptosis by the mitochondrial death pathway[149].

1.3.3.3. Suicide genes in development

The immunogenic nature of non-mammalian suicide genes such as HSV-TK and cytosine deaminase and the unintended immune mediated elimination of suicide gene expressing cells has prompted the search for novel human and/or non-

immunogenic genes able to function as suicide genes. A human T cell surface antigen, CD20, was one of the first human suicide genes to be investigated for its capacity to eliminate CD20 expressing T cells using anti-CD20 antibodies. The CD20/anti CD20 mAb may be suitable for use in gene modified HSCs but requires high cellular expression of CD20 antigen and may also deplete normal CD20-expressing[140]. Other systems that could be useful include the FK-506-binding protein (FKBP-FAS)/AP20187, AP1903 dimerization system that relies on the selective induction of apoptosis by expressing pro-apoptotic Fas-ligand molecules intracellularly, to be activated by non-toxic chemically induced dimerization of the FKBP-FAS molecules. Another notable non-immunogenic system (iCasp9) relies on activating apoptosis in selected cells by fusing the death domains of Caspase-9 with FKBP elements, which can be induced to dimerize and activate apoptosis[150]. This system is currently being evaluated in an ongoing clinical trial for graft *versus* host disease (cited by Lupo-Stanghellini MT, *et al*., 2010).

In summary, the incorporation of safety switches in the form of suicide genes to eliminate gene modified cells would be essential and beneficial features in future clinical gene therapy. Ongoing efforts to develop suicide genes with increased prodrug sensitivity and reduced toxicity, as well as exploring novel systems to selectively induce cell death ought to be helpful adjuncts to improving the biosafety of human gene therapy – currently mainly in clinical trials.

1.3.4. References

1. Wu C, Dunbar CE. Stem cell gene therapy: the risks of insertional mutagenesis and approaches to minimize genotoxicity. Frontiers of medicine. (2011) 5:356-371.

2. Conwell CC, Huang L. Recent advances in non-viral gene delivery. Adv Genet. (2005) 53:3-18.

3. Yant SR, Meuse L, Chiu W, Ivics Z, *et al*. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet. (2000) 25:35-41.

4. Walisko O, Schorn A, Rolfs F, Devaraj A, *et al*. Transcriptional activities of the Sleeping Beauty transposon and shielding its genetic cargo with insulators. Mol Ther. (2008) 16:359-369.

5. Zhu J, Park CW, Sjeklocha L, Kren BT, *et al*. High-level genomic integration, epigenetic changes, and expression of sleeping beauty transgene. Biochemistry. (2010) 49:1507-1521.

6. Geurts AM, Hackett CS, Bell JB, Bergemann TL, *et al*. Structure-based prediction of insertion-site preferences of transposons into chromosomes. Nucleic Acids Res. (2006) 34:2803-2811.

7. Huang X, Guo H, Tammana S, Jung YC, *et al*. Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. Mol Ther. (2010) 18:1803-1813.

8. Ortiz-Urda S, Lin Q, Yant S, Keene D, *et al*. Sustainable correction of junctional epidermolysis bullosa via transposon-mediated nonviral gene transfer. Gene Ther. (2003) 10:1099-1104.

9. Izsvák Z, Chuah MKL, Vandendriessche T, Ivics Z, *et al*. Efficient stable gene transfer into human cells by the Sleeping Beauty transposon vectors. Methods. (2009) 49:287-297.

10. Aronovich EL, Bell JB, Khan SA, Belur LR*, et al*. Systemic correction of storage disease in MPS I NOD/SCID mice using the sleeping beauty transposon system. Mol Ther. (2009) 17:1136-1144.

11. Collier LS, Carlson CM, Ravimohan S, Dupuy AJ*, et al*. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature. (2005) 436:272-276.

12. Ohlfest JR, Frandsen JL, Fritz S, Lobitz PD*, et al*. 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. (2005) 105:2691-2698.

13. Zayed H, Izsvák Z, Walisko O, Ivics Z*, et al*. Development of hyperactive sleeping beauty transposon vectors by mutational analysis. Mol Ther. (2004) 9:292-304.

14. Ivics Z, Hackett PB, Plasterk RH, Izsvák Z*, et al*. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell. (1997) 91:501-510.

15. Geurts AM, Yang Y, Clark KJ, Liu G*, et al*. Gene transfer into genomes of human cells by the sleeping beauty transposon system. Mol Ther. (2003) 8:108-117.

16. Mátés L, Chuah MKL, Belay E, Jerchow B*, et al*. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet. (2009) 41:753-761.

17. Xue X, Huang X, Nodland SE, Mátés L*, et al*. Stable gene transfer and expression in cord blood-derived CD34+ hematopoietic stem and progenitor cells by a hyperactive Sleeping Beauty transposon system. Blood. (2009) 114:1319-1330.

18. Wilber A, Linehan JL, Tian X, Woll PS*, et al*. Efficient and stable transgene expression in human embryonic stem cells using transposon-mediated gene transfer. Stem Cells. (2007) 25:2919-2927.

19. Yant SR, Huang Y, Akache B, Kay MA, *et al*. Site-directed transposon integration in human cells. Nucleic Acids Res. (2007) 35:e50.

20. Ivics Z, Katzer A, Stüwe EE, Fiedler D, *et al*. Targeted Sleeping Beauty transposition in human cells. Mol Ther. (2007) 15:1137-1144.

21. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, *et al*. Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor. Hum Gene Ther. (2006) 17:960-967.

22. Staunstrup NH, Moldt B, Mátés L, Villesen P, *et al*. Hybrid lentivirus-transposon vectors with a random integration profile in human cells. Mol Ther. (2009) 17:1205-1214.

23. Hackett PB, Largaespada DA, Cooper LJ. A transposon and transposase system for human application. (1). Mol Ther. (2010) 18:674-683.

24. Huang X, Haley K, Wong M, Guo H, *et al*. Unexpectedly high copy number of random integration but low frequency of persistent expression of the Sleeping Beauty transposase after trans delivery in primary human T cells. Hum Gene Ther. (2010) 21:1577-1590.

25. Ding S, Wu X, Li G, Han M, *et al*. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell. (2005) 122:473-483.

26. Woltjen K, Michael IP, Mohseni P, Desai R, *et al*. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature. (2009) 458:766-770.

27. Rad R, Rad L, Wang W, Cadinanos J, *et al*. PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. Science. (2010) 330:1104-1107.

28. Wu SCY, Meir YJJ, Coates CJ, Handler AM, *et al*. 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. (2006) 103:15008-15013.

29. Liang Q, Kong J, Stalker J, Bradley A*, et al*. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. Genesis. (2009) 47:404-408.

30. Yusa K, Zhou L, Li MA, Bradley A*, et al*. A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A. (2011) 108:1531-1536.

31. Grabundzija I, Irgang M, Mátés L, Belay E*, et al*. Comparative analysis of transposable element vector systems in human cells. Mol Ther. (2010) 18:1200-1209.

32. Thyagarajan B, Guimarães MJ, Groth AC, Calos MP*, et al*. Mammalian genomes contain active recombinase recognition sites. Gene. (2000) 244:47-54.

33. Buchholz F, Stewart AF. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. Nat Biotechnol. (2001) 19:1047-1052.

34. Shimshek DR, Kim J, Hübner MR, Spergel DJ*, et al*. Codon-improved Cre recombinase (iCre) expression in the mouse. Genesis. (2002) 32:19-26.

35. Schmidt EE, Taylor DS, Prigge JR, Barnett S*, et al*. Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. Proc Natl Acad Sci U S A. (2000) 97:13702-13707.

36. Loonstra A, Vooijs M, Beverloo HB, Allak BA*, et al*. Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. Proc Natl Acad Sci U S A. (2001) 98:9209-9214.

37. Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M*, et al*. Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. J Mol Biol. (2003) 326:65-76.

38. Buchholz F, Angrand PO, Stewart AF. Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol. (1998) 16:657-662.

39. Nehlsen K, Schucht R, da Gama-Norton L, Krömer W*, et al*. Recombinant protein expression by targeting pre-selected chromosomal loci. BMC Biotechnol. (2009) 9:100.

40. Monaco AP, Larin Z. YACs, BACs, PACs and MACs: artificial chromosomes as research tools. Trends Biotechnol. (1994) 12:280-286.

41. Yu Y, Bradley A. Engineering chromosomal rearrangements in mice. Nat Rev Genet. (2001) 2:780-790.

42. Wirth D, Gamanorton L, Riemer P, Sandhu U*, et al*. Road to precision: recombinase-based targeting technologies for genome engineering. Curr Opin Biotechnol. (2007) 18:411-419.

43. Groth AC, Olivares EC, Thyagarajan B, Calos MP*, et al*. A phage integrase directs efficient site-specific integration in human cells. Proc Natl Acad Sci U S A. (2000) 97:5995-6000.

44. Calos MP. The phiC31 integrase system for gene therapy. Curr Gene Ther. (2006) 6:633-645.

45. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B*, et al*. Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol. (2006) 357:28-48.

46. Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q*, et al*. Stable nonviral genetic correction of inherited human skin disease. Nat Med. (2002) 8:1166-1170.

47. Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q*, et al*. PhiC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. Hum Gene Ther. (2003) 14:923-928.

48. Thyagarajan B, Liu Y, Shin S, Lakshmipathy U*, et al*. Creation of engineered human embryonic stem cell lines using phiC31 integrase. Stem Cells. (2008) 26:119-126.

49. Ishikawa Y, Tanaka N, Murakami K, Uchiyama T, *et al*. Phage phiC31 integrase-mediated genomic integration of the common cytokine receptor gamma chain in human T-cell lines. J Gene Med. (2006) 8:646-653.

50. Held PK, Olivares EC, Aguilar CP, Finegold M, *et al*. In vivo correction of murine hereditary tyrosinemia type I by phiC31 integrase-mediated gene delivery. Mol Ther. (2005) 11:399-408.

51. Olivares EC, Hollis RP, Chalberg TW, Meuse L, *et al*. Site-specific genomic integration produces therapeutic Factor IX levels in mice. Nat Biotechnol. (2002) 20:1124-1128.

52. Bertoni C, Jarrahian S, Wheeler TM, Li Y, *et al*. Enhancement of plasmid-mediated gene therapy for muscular dystrophy by directed plasmid integration. Proc Natl Acad Sci U S A. (2006) 103:419-424.

53. Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, *et al*. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol. (2001) 21:3926-3934.

54. Liu J, Jeppesen I, Nielsen K, Jensen TG, *et al*. Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther. (2006) 13:1188-1190.

55. Ehrhardt A, Engler JA, Xu H, Cherry AM, *et al*. Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther. (2006) 17:1077-1094.

56. Sivalingam J, Krishnan S, Ng WH, Lee SS, *et al*. Biosafety assessment of site-directed transgene integration in human umbilical cord-lining cells. Mol Ther. (2010) 18:1346-1356.

57. Varella-Garcia M, Chen L, Powell RL, Hirsch FR, *et al*. Spectral karyotyping detects chromosome damage in bronchial cells of smokers and patients with cancer. Am J Respir Crit Care Med. (2007) 176:505-512.

58. Miller DG, Trobridge GD, Petek LM, Jacobs MA*, et al*. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J Virol. (2005) 79:11434-11442.

59. Sclimenti CR, Thyagarajan B, Calos MP. Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res. (2001) 29:5044-5051.

60. Liesner R, Zhang W, Noske N, Ehrhardt A*, et al*. Critical amino acid residues within the φC31 integrase DNA-binding domain affect recombination activities in mammalian cells. Hum Gene Ther. (2010) 21:1104-1118.

61. Keravala A, Lee S, Thyagarajan B, Olivares EC*, et al*. Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. Mol Ther. (2009) 17:112-120.

62. Izsvák Z, Hackett PB, Cooper LJ, Ivics Z*, et al*. Translating Sleeping Beauty transposition into cellular therapies: victories and challenges. Bioessays. (2010) 32:756-767.

63. Jang MY, Yarborough OH, Conyers GB, McPhie P*, et al*. Stable secondary structure near the nicking site for adeno-associated virus type 2 Rep proteins on human chromosome 19. J Virol. (2005) 79:3544-3556.

64. Pieroni L, Fipaldini C, Monciotti A, Cimini D*, et al*. Targeted integration of adeno-associated virus-derived plasmids in transfected human cells. Virology. (1998) 249:249-259.

65. Philpott NJ, Gomos J, Berns KI, Falck-Pedersen E*, et al*. A p5 integration efficiency element mediates Rep-dependent integration into AAVS1 at chromosome 19. Proc Natl Acad Sci U S A. (2002) 99:12381-12385.

66. Feng D, Chen J, Yue Y, Zhu H*, et al*. A 16bp Rep binding element is sufficient for mediating Rep-dependent integration into AAVS1. J Mol Biol. (2006) 358:38-45.

67. Howden SE, Voullaire L, Vadolas J. The transient expression of mRNA coding for Rep protein from AAV facilitates targeted plasmid integration. J Gene Med. (2008) 10:42-50.

68. Tsunoda H, Hayakawa T, Sakuragawa N, Koyama H, *et al*. Site-specific integration of adeno-associated virus-based plasmid vectors in lipofected HeLa cells. Virology. (2000) 268:391-401.

69. Liu R, Li Y, Hu R, Jin T, *et al*. A site-specific genomic integration strategy for sustained expression of glucagon-like peptide-1 in mouse muscle for controlling energy homeostasis. Biochem Biophys Res Commun. (2010) 403:172-177.

70. Katz RA, Merkel G, Skalka AM. Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. Virology. (1996) 217:178-190.

71. Bushman FD. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc Natl Acad Sci U S A. (1994) 91:9233-9237.

72. Bushman FD, Miller MD. Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. J Virol. (1997) 71:458-464.

73. Gijsbers R, Ronen K, Vets S, Malani N, *et al*. LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol Ther. (2010) 18:552-560.

74. Venter JC, Adams MD, Myers EW, Li PW, *et al*. The sequence of the human genome. Science. (2001) 291:1304-1351.

75. Tan W, Dong Z, Wilkinson TA, Barbas CF, *et al*. Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. J Virol. (2006) 80:1939-1948.

76. Rouet P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc Natl Acad Sci U S A. (1994) 91:6064-6068.

77. Arnould S, Perez C, Cabaniols JP, Smith J, *et al*. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol. (2007) 371:49-65.

78. Grizot S, Smith J, Daboussi F, Prieto J, *et al*. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res. (2009) 37:5405-5419.

79. Fajardo-Sanchez E, Stricher F, Pâques F, Isalan M, *et al*. Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. Nucleic Acids Res. (2008) 36:2163-2173.

80. Li H, Pellenz S, Ulge U, Stoddard BL, *et al*. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. Nucleic Acids Res. (2009) 37:1650-1662.

81. Ashworth J, Havranek JJ, Duarte CM, Sussman D, *et al*. Computational redesign of endonuclease DNA binding and cleavage specificity. Nature. (2006) 441:656-659.

82. Galetto R, Duchateau P, Pâques F. Targeted approaches for gene therapy and the emergence of engineered meganucleases. Expert Opin Biol Ther. (2009) 9:1289-1303.

83. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A. (1996) 93:1156-1160.

84. Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. Nat Biotechnol. (2005) 23:967-973.

85. Miller JC, Holmes MC, Wang J, Guschin DY, *et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. (2007) 25:778-785.

86. Szczepek M, Brondani V, Büchel J, Serrano L, *et al*. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol. (2007) 25:786-793.

87. Segal DJ, Beerli RR, Blancafort P, Dreier B, *et al*. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry. (2003) 42:2137-2148.

88. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, *et al*. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell. (2008) 31:294-301.

89. Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, *et al*. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods. (2011) 8:67-69.

90. Delacôte F, Lopez BS. Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model. Cell Cycle. (2008) 7:33-38.

91. Hartlerode AJ, Scully R. Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J. (2009) 423:157-168.

92. Barzel A, Kupiec M. Finding a match: how do homologous sequences get together for recombination? Nat Rev Genet. (2008) 9:27-37.

93. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. Cell Res. (2008) 18:134-147.

94. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, *et al*. Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics. (2010) 11:636-646.

95. Shukla VK, Doyon Y, Miller JC, DeKelver RC, *et al*. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature. (2009) 459:437-441.

96. Townsend JA, Wright DA, Winfrey RJ, Fu F, *et al*. High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature. (2009) 459:442-445.

97. Rémy S, Tesson L, Ménoret S, Usal C, *et al*. Zinc-finger nucleases: a powerful tool for genetic engineering of animals. Transgenic Res. (2010) 19:363-371.

98. Hockemeyer D, Soldner F, Beard C, Gao Q, *et al*. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. (2009) 27:851-857.

99. McCammon JM, Amacher SL. Using zinc finger nucleases for efficient and heritable gene disruption in zebrafish. Methods Mol Biol. (2010) 649:281-298.

100. Carroll D, Beumer KJ, Trautman JK. High-efficiency gene targeting in Drosophila with zinc finger nucleases. Methods Mol Biol. (2010) 649:271-280.

101. Perez EE, Wang J, Miller JC, Jouvenot Y, *et al*. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol. (2008) 26:808-816.

102. Cost GJ, Freyvert Y, Vafiadis A, Santiago Y, *et al*. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol Bioeng. (2010) 105:330-340.

103. Liu PQ, Chan EM, Cost GJ, Zhang L, *et al*. Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. Biotechnol Bioeng. (2010) 106:97-105.

104. Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. Science. (2003) 300:763.

105. Urnov FD, Miller JC, Lee YL, Beausejour CM, *et al*. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature. (2005) 435:646-651.

106. Moehle EA, Rock JM, Lee YL, Jouvenot Y, *et al*. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proceedings of the National Academy of Sciences. (2007) 104:3055-3060.

107. Benabdallah BF, Allard E, Yao S, Friedman G, *et al*. Targeted gene addition to human mesenchymal stromal cells as a cell-based plasma-soluble protein delivery platform. Cytotherapy. (2010) 12:394-399.

108. Lombardo A, Genovese P, Beausejour CM, Colleoni S, *et al*. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol. (2007) 25:1298-1306.

109. Sebastiano V, Maeder ML, Angstman JF, Haddad B, *et al*. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. Stem Cells. (2011) 29:1717-1726.

110. Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, *et al*. Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Res. (2009) 37:D279-D283.

111. Sander JD, Maeder ML, Reyon D, Voytas DF, *et al*. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. (2010) 38:W462-W468.

112. Reyon D, Kirkpatrick JR, Sander JD, Zhang F, *et al*. ZFNGenome: A comprehensive resource for locating zinc finger nuclease target sites in model organisms. BMC Genomics. (2011) 12:83.

113. Cradick TJ, Ambrosini G, Iseli C, Bucher P, *et al*. ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. BMC Bioinformatics. (2011) 12:152.

114. Guo J, Gaj T, Barbas CF. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol. (2010) 400:96-107.

115. Doyon Y, Vo TD, Mendel MC, Greenberg SG*, et al*. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. (2011) 8:74-79.

116. Doyon Y, Choi VM, Xia DF, Vo TD*, et al*. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat Methods. (2010) 7:459-460.

117. Ramirez CL, Certo MT, Mussolino C, Goodwin MJ*, et al*. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucleic Acids Res. (2012) 40:5560-5568.

118. Wang J, Friedman G, Doyon Y, Wang NS*, et al*. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. Genome Res. (2012) 22:1316-1326.

119. Mussolino C, Cathomen T. TALE nucleases: tailored genome engineering made easy. Curr Opin Biotechnol. (2012) 23:644-650.

120. Boch J, Scholze H, Schornack S, Landgraf A*, et al*. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. (2009) 326:1509-1512.

121. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science. (2009) 326:1501.

122. Christian M, Cermak T, Doyle EL, Schmidt C*, et al*. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. (2010) 186:757-761.

123. Weber E, Gruetzner R, Werner S, Engler C*, et al*. Assembly of designer TAL effectors by Golden Gate cloning. PloS one. (2011) 6:e19722.

124. Reyon D, Tsai SQ, Khayter C, Foden JA*, et al*. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol. (2012) 30:460-465.

125. Wu J, Kandavelou K, Chandrasegaran S. Custom-designed zinc finger nucleases: what is next? Cell Mol Life Sci. (2007) 64:2933-2944.

126. Mussolino C, Morbitzer R, Lütge F, Dannemann N, *et al*. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. (2011) 39:9283-9293.

127. Sun N, Abil Z, Zhao H. Recent advances in targeted genome engineering in mammalian systems. Biotechnol J. (2012) 7:1074-1087.

128. Sadelain M, Papapetrou EP, Bushman FD. Safe harbours for the integration of new DNA in the human genome. Nat Rev Cancer. (2011) 12:51-58.

129. Liu X, Wu Y, Li Z, Yang J, *et al*. Targeting of the human coagulation factor IX gene at rDNA locus of human embryonic stem cells. PloS one. (2012) 7:e37071.

130. Irion S, Luche H, Gadue P, Fehling HJ, *et al*. Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat Biotechnol. (2007) 25:1477-1482.

131. Costa M, Dottori M, Ng E, Hawes SM, *et al*. The hESC line Envy expresses high levels of GFP in all differentiated progeny. Nat Methods. (2005) 2:259-260.

132. Sakurai K, Shimoji M, Tahimic CGT, Aiba K, *et al*. Efficient integration of transgenes into a defined locus in human embryonic stem cells. Nucleic Acids Res. (2010) 38:e96.

133. Mulder J, Ariaens A, van den Boomen D, Moolenaar WH, *et al*. p116Rip targets myosin phosphatase to the actin cytoskeleton and is essential for RhoA/ROCK-regulated neuritogenesis. Mol Biol Cell. (2004) 15:5516-5527.

134. Smith RH. Adeno-associated virus integration: virus versus vector. Gene Ther. (2008) 15:817-822.

135. Smith JR, Maguire S, Davis LA, Alexander M, *et al*. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. Stem Cells. (2008) 26:496-504.

136. Ogata T, Kozuka T, Kanda T. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. J Virol. (2003) 77:9000-9007.

137. Zou C, Chou BK, Dowey SN, Tsang K, *et al*. Efficient derivation and genetic modifications of human pluripotent stem cells on engineered human feeder cell lines. Stem Cells Dev. (2012) 21:2298-2311.

138. Chang CJ, Bouhassira EE. Zinc-finger nuclease mediated correction of α-thalassemia in iPS cells. Blood. (2012)

139. Denny WA. Prodrugs for Gene-Directed Enzyme-Prodrug Therapy (Suicide Gene Therapy). Journal of biomedicine & biotechnology. (2003) 2003:48-70.

140. Lupo-Stanghellini M, Provasi E, Bondanza A, Ciceri F, *et al*. Clinical impact of suicide gene therapy in allogeneic hematopoietic stem cell transplantation. Hum Gene Ther. (2010) 21:241-250.

141. Garin MI, Garrett E, Tiberghien P, Apperley JF, *et al*. Molecular mechanism for ganciclovir resistance in human T lymphocytes transduced with retroviral vectors carrying the herpes simplex virus thymidine kinase gene. Blood. (2001) 97:122-129.

142. Balzarini J, Liekens S, Solaroli N, El Omari K, *et al*. Engineering of a single conserved amino acid residue of herpes simplex virus type 1 thymidine kinase allows a predominant shift from pyrimidine to purine nucleoside phosphorylation. J Biol Chem. (2006) 281:19273-19279.

143. Berger C, Flowers ME, Warren EH, Riddell SR, *et al*. Analysis of transgene-specific immune responses that limit the in vivo persistence of adoptively transferred HSV-TK-modified donor T cells after allogeneic hematopoietic cell transplantation. Blood. (2006) 107:2294-2302.

144. Chalmers D, Ferrand C, Apperley JF, Melo JV, *et al*. Elimination of the truncated message from the herpes simplex virus thymidine kinase suicide gene. Mol Ther. (2001) 4:146-148.

145. Black ME, Kokoris MS, Sabo P. Herpes simplex virus-1 thymidine kinase mutants created by semi-random sequence mutagenesis improve prodrug-mediated tumor cell killing. Cancer Res. (2001) 61:3022-3026.

146. Mercer KE, Ahn CE, Coke A, Compadre CM, *et al*. Mutation of herpesvirus thymidine kinase to generate ganciclovir-specific kinases for use in cancer gene therapies. Protein Eng. (2002) 15:903-911.

147. Preuß E, Treschow A, Newrzela S, Brücher D, *et al*. TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. Hum Gene Ther. (2010) 21:929-941.

148. Kuriyama S, Masui K, Sakamoto T, Nakatani T, *et al*. Bystander effect caused by cytosine deaminase gene and 5-fluorocytosine in vitro is substantially mediated by generated 5-fluorouracil. Anticancer Res. (1998) 18:3399-3406.

149. Sato T, Neschadim A, Konrad M, Fowler DH, *et al*. Engineered human tmpk/AZT as a novel enzyme/prodrug axis for suicide gene therapy. Mol Ther. (2007) 15:962-970.

150. Tey SK, Dotti G, Rooney CM, Heslop HE, *et al*. Inducible caspase 9 suicide gene to improve the safety of allodepleted T cells after haploidentical stem cell transplantation. Biol Blood Marrow Transplant. (2007) 13:913-924.

1.4.   Hemophilia A as a model disease for gene therapy

1.4.1.   Coagulation pathway and bleeding disorders

Evolution has equipped all mammals with a conserved mechanism for maintaining hemostasis and evading life threatening bleeds. Coagulation, the process by which blood turns into an insoluble clot of fibrinogen, serves as the secondary hemostasis mechanism *in vivo*, immediately following primary hemostasis provided by platelets that are recruited to and form a plug at the damaged endothelium of blood vessels. The coagulation cascade refers to the series of biochemical events that culminate in fibrin clot formation. It involves the near-instantaneous and sequential activation of several zymogen proteases (inactive precursors) and their glycoprotein cofactors which in turn catalyze the activation of other proteins (clotting factors) that are essential in the eventual formation of fibrin polymers that strengthen platelet plugs at sites of vascular injury.

Coagulation can be broadly separated into two pathways, the primary pathway being the tissue factor or extrinsic pathway[1] and the other being the contact dependent or intrinsic pathway[2]. These two pathways converge on a final common pathway that involves factor X (FX), thrombin, factor XIII (FXIII) and fibrinogen, culminating in the polymerization and cross-linking of fibrin monomers to form a clot (**Figure 1.4.1**). In the extrinsic pathway, damage to blood vessels or extravascular tissues exposes an integral membrane protein, called the tissue factor (TF), to binding by a circulating plasma protein, factor VII (FVII), thereby activating the latter and initiating the coagulation process. Activated FVII (FVIIa), in the presence of calcium ions and phospholipids, activates FX to FXa which, together with activated factor Va (FVa), cleaves inactive prothrombin to form activated thrombin. Activated thrombin in turn cleaves fibrinogen into fibrin monomers which are covalently polymerized by activated FXIII (FXIIIa) to form a fibrin clot. Thrombin is also responsible for the activation of FV, FVIII, FXI, FXIII and another protein, protein C, which is involved in the down-regulation of the coagulation cascade. Activated FVIII (FVIIIa) and FIX (FIXa) are two essential clotting factors that serve to greatly amplify the activation of FXa and thus the coagulation process. Whilst the tissue factor pathway is initiated by the presence of tissue factors from tissue injury, the intrinsic pathway is activated by contact of blood with surfaces such as collagen on damaged vascular wall or any charged or wettable surfaces such as glass[2]. The binding of high molecular weight kininogen (HMWK) and prekallikrein to collagen, initiates the activation of FXII to FXIIa which kick starts the coagulation cascade by activating FXI to FXIa, in turn converting FIX to FIXa. The contact activation pathway converges with the tissue

factor pathway with the activation of FX to FXa by FVIIIa and FIXa, culminating in formation of a fibrin clot. The process by which fibrin clots are eventually broken down and resorbed is known as fibrinolysis and this is controlled mainly by another plasma protein known as plasmin.
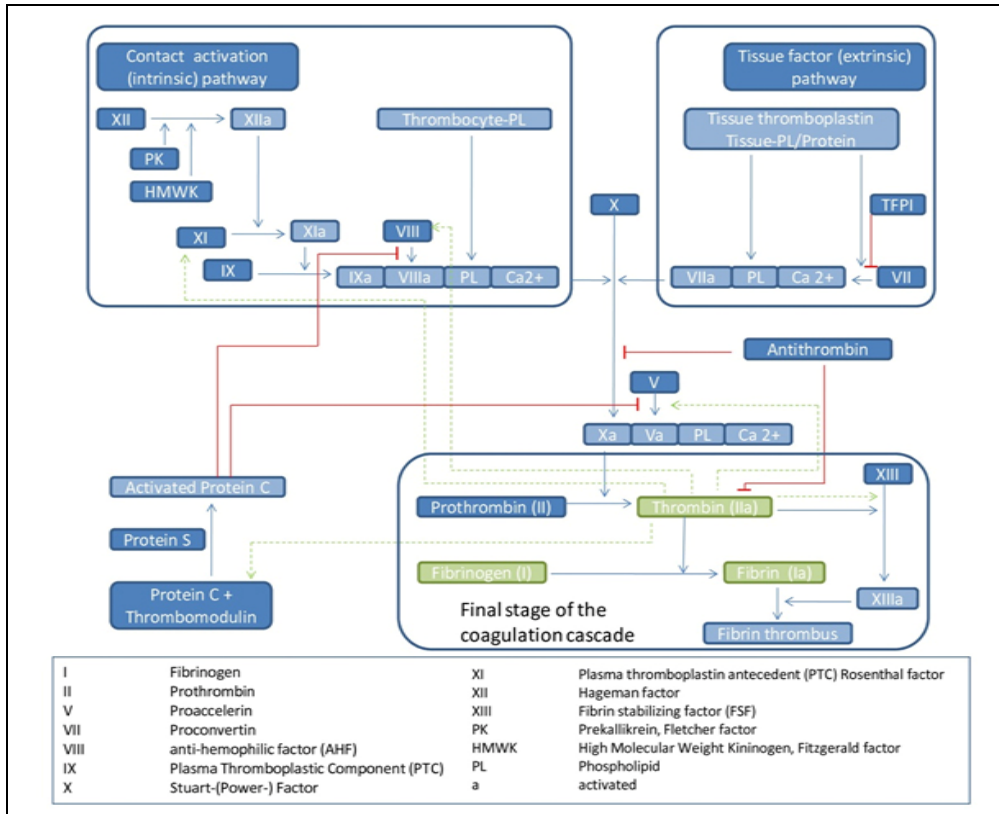
Figure 1.4.1   Coagulation cascade. Coagulation is activated either by contact activation of blood with surfaces such as collagen on damaged vascular walls (intrinsic pathway) or by binding of tissue factor (TF) from injured blood vessels (extrinsic pathway) to circulating plasma protein, factor VII. In the intrinsic pathway, activated FIXa and FVIIIa in association with phospholipids and divalent calcium ions (tenase complex) serve to activate a key constituent of the clotting cascade, FX. In the extrinsic pathway, activated FVIIa is responsible for the activation of FX. Activated FXa together with FVa, catalyzes conversion of prothrombin to thrombin. Thrombin is necessary for amplification of other components of the coagulation cascade (green dotted lines) and for breakdown of fibrinogen to fibrin monomers which are polymerized by activated FXIIIa to form a fibrin thrombus clot. Negative regulators of the coagulation cascade (red lines) include activated protein C and antithrombin. (*Figure taken from ref. 3.*)[3]

Given the involvement of several clotting factors, it is easy to appreciate that deficiencies in key clotting factors will perturb normal hemostasis and cause bleeding disorders of varying phenotypic severities. The most common bleeding disorders due to coagulation factor deficiencies are the hemophilias A and B which are caused by deficiencies in clotting factors FVIII and FIX, respectively. FXI deficiency is very rare and results in a milder bleeding disorder characterized by trauma and soft tissue-

related hemorrhage[4] while FXII deficiency is not associated with clinically significant bleeding diathesis[5].

1.4.2. Brief history of hemophilia A

Hereditary hemophilia A is an X-chromosome linked recessive monogenic bleeding disorder affecting 1 in 5000 -10000 males[6]. This genetic disorder, arising from mutations in FVIII gene, results in impaired coagulation due to either a deficient or dysfunctional FVIII protein and is characterized by sporadic, uncontrolled and unarrested bleeding[7]. Although males are primarily affected, female heterozygous carriers bearing a defective FVIII gene may also manifest clinical FVIII deficiency depending on other factors such as inactivation of the normal X chromosome and maternal or paternal inheritance of the defective X chromosome. Acquired hemophilia A is less common, occurs mainly among the elderly and is caused by autoinhibitory antibodies against FVIII protein rather than FVIII gene mutations[8].

Cases of familial bleeding disorders were recorded in Jewish rabbinical writings as early as the 2nd century AD when male babies were spared circumcision if there had been previous deaths in the family from this procedure. One of the first modern day descriptions was an 1803 publication by John Conrad Otto entitled "An account of a hemorrhagic disposition existing in certain families."(Otto, JC *et al*. 1803) The term hemophilia itself was first used to describe a bleeding disorder at the University of Zurich in 1828[9]. The best known cases of hemophilia have been traced to the royal families in Europe. Queen Victoria of England (1837 – 1901), was historically thought to have harbored mutations in the FVIII gene, that was eventually passed on to several of her royal descendents, eventually affecting several royal families throughout Europe – most notably the family of Nicholas II, the last Czar of Russia, with tragic consequences[7]. However, new evidence shows that the 'royal disease' was more likely due to mutations in the FIX gene rather than the FVIII gene. As early as the 1940s, scientists *via* blood transfusion studies linked this bleeding disorder to a deficiency of a "certain factor" in the blood. In 1937, Harvard University doctors Patek and Taylor identified a plasma extract, they termed anti-hemophilic globulin, that was capable of correcting coagulation defects[10]. Thus, up till the 1950s and 1960s, hemophilia A and other bleeding disorders were treated with whole blood or fresh plasma infusions. It was not until the mid-1960s that FVIII was purified as a highly concentrated cryoprecipitate which enabled a more practical and efficient form of FVIII replacement therapy[11].

1.4.3. Anti-hemophilic factor, FVIII

FVIII is a plasma protein synthesized mainly in the liver by hepatocytes and hepatic sinusoidal endothelial cells. Infused FVIII has a mean physiological half-life of 12 hours in plasma where it is stabilized and protected from proteolytic inactivation by its non-covalent association with vonWillebrand factor (VWF), another plasma protein secreted by endothelial cells and megakaryocytes[12]. Activated FVIII (FVIIIa) together with activated FIX (FIXa) serves as a cofactor for activation of FX to Xa in the intrinsic (contact activation) coagulation pathway. Lack or absence of functional FVIII protein results in a clinical bleeding disorder of varying severity. Severity of the bleeding diathesis correlates with levels of plasma FVIII activity. Thus, hemophilia A is classified as mild (5-40 % FVIII activity), moderate (1-5% FVIII activity) or severe (<1 % FVIII activity)[13]. Individuals with hemophilia A often bleed spontaneously, especially into gums, large joints such as the knees, ankles and elbows, and less often but with greater morbidity into the hips, muscles, intra-abdominal and intra-cranial cavities. Bleeding into the latter sites can be life threatening and fatal. Complications arising from chronic bleeding include pain, numbness, hemophilic pseudo-tumors (blood cysts), hemarthrosis (joint bleed), hemophilic arthropathy (joint damage and disfigurement), muscle wasting and crippling deformities and debilitating arthritis[13].

The FVIII gene which spans a 186 kb genomic region in Xq28 was successfully cloned and sequenced in 1984[14, 15]. It comprises 26 coding exons that are transcribed to yield a 9-kb mRNA which is translated to a glycosylated precursor polypeptide of 2351 amino acids having a signal peptide of 19 amino acids. The mature FVIII protein consisting of 2332 amino acids has the domain structure, from N- to C-terminus, of A1-a1-A2-a2-B-a3-A3-C1-C2. It is secreted and circulates as a heterodimer bound to VWF comprising A1-A2-B heavy chain (90-210 kD) non-covalently bound to the A3-C1-C2 light chain (80 kD). Circulating FVIII is activated by thrombin which catalyzes limited proteolysis to form heterotrimeric FVIIIa consisting of A1, A2 and A3-C1-C2 subunits [16]. The B-domain, which is involved in intracellular processing of the primary FVIII protein, is functionally dispensable for blood coagulation. It is thus not a component of activated FVIII protein in plasma.
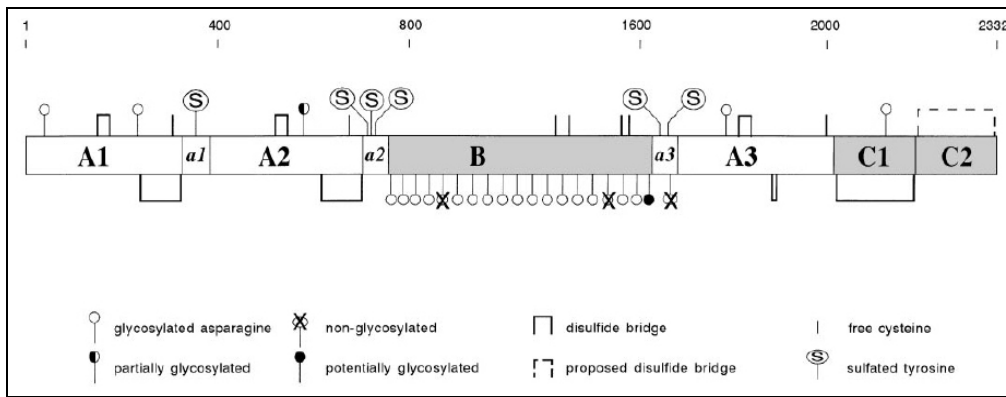
Figure 1.4.2   Structure of FVIII protein. FVIII protein consist of domains A1 (residues 1-336), A2 (373-710), B (741-1648), A3 (1690-2019), C1 (2020-2172) and C2 (2173-2332) which are separated by acidic regions (a1,a2 and a3). Presence of predicted and know disulphide bridges, free cysteine-residues, sulphated tyrosine and N-linked glycosylation sites within the FVIII protein are indicated. (*Figure source; ref. 16*)

FVIIIa dissociates from VWF and interacts with FIXa, phospholipid and FX to convert FX to FXa, thereby activating the clotting cascade. FVIII catabolism, its breakdown and removal from circulation, is thought to involve hepatocytes through binding to endocytic receptors such as low-density lipoprotein receptor-related protein (LRP) and low-density lipoprotein receptor (LDLR)[17]. This is facilitated by binding to cell surface heparan sulphate proteoglycans (HSPG), glycoprotein components of the extracellular matrix, which serve to concentrate FVIII on the cell surface for LRP mediated clearance[18]. Plasma FVIII levels are also thought to be regulated by binding to other members of low-density lipoprotein (LDL) receptor superfamily such as megalin which are expressed in kidney and very-low-density lipoprotein (VLDL) which are expressed in endothelial cells[19].

1.4.4.   Genetic mutations and FVIII deficiency

FVIII deficiency or dysfunction is caused by mutations of the FVIII gene which result in significantly reduced production of active FVIII protein, production of an inactive (often truncated) form of FVIII protein or a total lack of functional FVIII protein. Owing in part to its large size, the FVIII gene has a high mutation rate[20] ranging from 2.5×10^{-5} to 4.2×10^{-5}. These mutations are approximately 3.6 times more prevalent in male compared to female germ cells. Amongst these genetic abnormalities, inversion of intron 22 is most prevalent, occurring in as many as one-third of patients. This is followed by missense mutations which account for 38% of cases, small deletions and insertions in 10%, nonsense mutations in 9%, while about one-third of patients have novel mutations. This diversity of gene abnormalities is

reflected in the phenotypic variations of functional defects of mutated FVIII gene products. In addition to the common intron 22 inversion, 40% of point mutations usually occur in one of 70 CpG sites within the FVIII gene and about 25% of small deletions are concentrated within exon 14[20]. A comprehensive catalog of all FVIII mutations can be found in the Hemophilia A Mutation, Structure, Test and Resourse Site (HAMSTeRS: http://hadb.org.uk). In rare cases, hemophilic phenotypes can also arise from mutations in genes other than FVIII. For instance, mutations affecting FVIII binding sites on VWF give rise to type 2N von Willebrand disease (VWD)[21], characterized by a severely decreased half-life of circulating FVIII. Another phenotype may be caused by combined deficiency of FV and FVIII (F5F8D)[22] as a result of mutations to *LMAN1* and *MCFD2*. The latter two genes encode proteins involved in the secretory pathway and cause defective secretion of FV and FVIII when mutated.

1.4.5.   Treatment options for hemophilia A
1.4.5.1. Early history to present day treatment

As early as the 1940s, scientists and physicians had already attributed bleeding disorders to deficiencies of certain factors in blood. Replacement of these deficient factors in patients with hemophilia using whole blood or fresh plasma from normal patients was the common treatment until the 1950s to 1960s[9]. However, owing to the low quantities of coagulation factors present, treatment required infusion of impractically large volumes of blood or plasma in order to arrest a serious bleeding episode.

It was not until the production of FVIII as a highly concentrated cryoprecipitate from plasma by Judith Pool in 1964 that a more practical and effective FVIII replacement therapy was introduced[11]. By the 1970s, the availability of lyophilized FVIII cryoprecipitates prepared from pooled donor plasma samples radically improved hemophilia treatment and made home FVIII replacement procedures feasible. Regular FVIII administration pioneered by Inge Marie Nilsson and Ake Ahlberg[23] improved severe hemophilia to a milder form, and aimed to prevent rather than only treat serious bleeding episodes. For the first time, these advances gave patients options to elect for surgical procedures such as for correction of musculoskeletal complications.

The golden era of hemophilia treatment suffered a major setback during the early 1980s, when many hemophiliacs were found to have been infected with viruses that contaminated plasma-derived FVIII products, specifically hepatitis B and C, and

HIV[24]. Lack of donor screening and without methods to eradicate or inactive viruses had inevitably resulted in FVIII concentrates prepared from infected donor samples, casting a pall on hemophilia treatment. Since then, the safety of plasma-derived FVIII concentrates has improved markedly from stringent screening of blood donors for exposure to viruses and equally, from technical advances in methods for inactivating viruses and other pathogens in plasma-derived products. This is evident from the fact that there have been no new cases of transmitted hepatitis virus or HIV infection from plasma-derived FVIII concentrates in the past 25 years [23]. However, eradication of other blood-borne pathogens, such as B19 parvovirus[25] and transmissible prion protein causing new variant Creutzfeldt-Jakob disease (nvCJD) remains at least a theoretical concern[26]. Though such risks are probably low due to improved purification procedures, they have not been completely abolished.

Advances in recombinant DNA technology and cloning of the FVIII gene in 1982[14, 15] proved to be major steps forward for hemophilia treatment as they paved the way for large scale production of recombinant FVIII, free from the risks of blood-borne pathogens but are extremely costly.

The advent of safe FVIII concentrates has thus enabled life-long treatment options for people with hemophilia who now benefit from improved life expectancy that is predicted to be only 10 years shorter than normal unaffected males (i.e. 50-60 years). This is a significant improvement considering that the mean life expectancy was only 11.4 years prior to the availability of effective treatments[27]. Current treatment options available for hemophilia care will be discussed in greater detail below. However, there remain major challenges for hemophilia care that need to be resolved. Up to 25% of patients with hemophilia A develop neutralizing or inhibitory antibodies against FVIII that make replacement therapies ineffective[28]. There is currently no consistently effective method to induce immune tolerance to FVIII protein or to make FVIII less immunogenic. Although hemophilia care has greatly improved over the decades, two-thirds of patients live in developing countries and most have poor or no access to proper hemophilia care[29]. This powerfully motivates development of alternative options such as gene therapy that could potentially offer long term and affordable treatment for people with hemophilia throughout the world.

1.4.5.2. FVIII replacement therapy

Currently, FVIII replacement therapy via intravenous infusion of plasma derived or recombinant FVIII concentrates is the mainstay of hemophilia A treatment. Cost aside, the dose and frequency of FVIII replacement is determined by the severity

of the disease. FVIII infusions may be administered prophylactically to prevent bleeding as well as on-demand (episodic) to treat serious bleeding episodes by trained professionals at hospitals, specialized clinics and comprehensive treatment centers or even by trained individuals at home. There is now evidence to support the view that prophylactic FVIII replacement from an early age, if affordable, should be the standard of clinical care[30]. Based on pharmacokinetic studies it has generally been established that there is a 2 -2.5% increase in FVIII activity for every unit of FVIII per kg body weight infused, thus establishing a factor of 0.5 IU/kg body weight[31]. The dosage of FVIII required to achieve a desired increase in plasma FVIII level thus can be calculated by the general formula:

Dose of FVIII required (IU) = Desired increase in FVIII (%) x *Body weight (kg)* x *0.5 (IU/kg)*

A more detailed formula to calculate the steady state concentration of plasma FVIII at any given time following a single infusion of FVIII concentrate is the following[32]:

FVIII (t) = Dose x *IVR* x $e^{-k \times (t-1)}$

where FVIII (t) is the concentration of plasma FVIII (IU) at any given time (t) in hours post-infusion, Dose is the concentration of FVIII infused per kg body weight (IU kg^{-1}), IVR is the measured *in vivo* recovery (IU dL^{-1} per IU kg^{-1}) and k is the elimination rate constant (In2/half-life).

Prophylactic treatment entails infusion of FVIII concentrate every 2 to 3 days to ensure trough plasma FVIII are maintained above 1% (1 IU dL^{-1}). The Medical and Scientific Advisory Committee of the National Hemophilia Foundation[33] (http://www.hemophilia.org/NHFWeb/MainPgs/MainNHF.aspx?menuid=57&contentid=1007, Accessed November 2012) and the World Health Organization[34] recommend prophylactic treatment for severe hemophilia. Several studies have shown the benefits of prophylactic FVIII treatment [35] to include prevention or decreased frequency of sporadic bleeding episodes, reduced pain, reduced risk of developing debilitating hemophilic arthopathy and improved quality of life enabling patients to participate in normal activities of daily living including various sports. The observed beneficial effects correlate with compliance to prophylactic regimes and maintenance of trough levels of plasma FVIII.

However at current prices of recombinant FVIII and other high quality FVIII products, the cost of prophylactic treatment is extremely high, viz. in the range of US$100, 000 per annum for plasma-derived FVIII concentrates to US$300,000 per annum for recombinant FVIII [35]. Furthermore, lack of quality FVIII concentrates in some countries restricts patients to episodic FVIII infusions for acute bleeds rather

than to prevent them. The goal of on-demand FVIII replacement treatment is to increase plasma FVIII levels to 50% in order to arrest hemorrhage, the dose required will vary depending on the extent and location of the bleed.

1.4.5.2.1.   Plasma-derived FVIII concentrates

Plasma-derived factor concentrates are purified from pooled plasma and are usually provided as lyophilized FVIII/vWF coprecipitates, stabilized with albumin or sucrose. Their purity depends on the methods used such as conventional cold ethanol precipitation, ion exchange chromatography, heparin affinity or immunoaffinity purification, of which the latter two methods yield higher purity concentrates.

The safety of plasma-derived factor concentrates has improved quite significantly over the years. Much of the initial hazards of iatrogenically transmitted viral infections from contaminated donor plasma samples have been effectively eliminated. Nucleic acid testing and other PCR-based methods are routinely employed to screen and quarantine plasma samples thought to carry pathogenic elements[36]. The use of dry/wet heating, solvents/detergents and combinations of virucidal methods (low pH treatment, sodium thiocyanate treatment) ensure the elimination and inactivation of HIV, hepatitis C and other non-enveloped viruses, if any present, in donor plasma samples[37]. Improved fractionation techniques such as selective precipitation and size exclusion chromatography used to purify and concentrate FVIII also minimize the risks of co-purifying contaminating prion proteins, thereby reducing concerns of transmissible encephalopathies from blood-borne infectious prions. In summary the current state of plasma-derived FVIII concentrates ensures an unparalleled level of safety compared with earlier products, but in no way ensures complete protection from other transmissible pathogenic agents.

Owing to the high cost and limited availability of recombinant FVIII, the use of plasma-derived FVIII concentrates still dominates as much as 20-30% of the FVIII replacement therapy market, even in affluent and developed countries[36]. Two other important but debatable factors favoring the use of plasma-derived FVIII are the lower incidence of FVIII inhibitor formation and higher rate of tolerance induction with the use of plasma-derived compared to recombinant FVIII[38].

1.4.5.2.2.   Recombinant FVIII concentrates

In 1984, four back-to-back articles in Nature reported the sequence of the human FVIII gene[39], protein structure of FVIII[14], cloning of its full length cDNA[15]

and demonstrated the capacity of recombinant FVIII protein to correct clotting deficiency in FVIII-deficient plasma[40]. These groundbreaking discoveries allowed for the first time an opportunity to treat patients with hemophilia A without having to rely on donor plasma. This became a reality in 1989 when clinical efficacy was shown in two patients with severe hemophilia treated with recombinant FVIII[41]. Soon thereafter, the industrial manufacture of recombinant FVIII starting from the early 1990s ushered in a new era of potentially safer hemophilia treatment.

Recombinant FVIII is prepared from conditioned culture media of transgenic mammalian cells (usually baby hamster kidney cells or Chinese hamster ovary cells) genetically modified to secrete either full-length or B-domain deleted FVIII. The FVIII protein is usually purified by chromatography such as ion-exchange, gel filtration and immunoaffinity purification, undergoes viral inactivation and is concentrated, lyophilized and formulated with stabilizers such as human albumin or, more recently, sucrose[42]. Over the years, the manufacture and purification of recombinant FVIII has significantly improved owing to improved protein purification and virucidal techniques, and avoidance of all human or animal proteins during the manufacture processes.

Several recombinant FVIII products are now marketed, differing in the nature of the FVIII protein (full-length or B domain-deleted), method of production (cell lines used; CHO or BHK), purification, viral inactivation and eventual formulation (stabilizers used such as albumin, sucrose, trehalose). Unlike most plasma-derived FVIII concentrates (except those purified with monoclonal antibodies), most recombinant FVIII products lack VWF. Although debatable[43], it is thought that VWF binding to FVIII may reduce its immunogenicity *via* epitope masking and may protect it from endocytosis by antigen-presenting cells thereby contributing to an overall lower immunogenicity of FVIII and lower risk of alloantibody formation against FVIII[44].

The use of recombinant FVIII is more prevalent in affluent and developed countries especially in Western European and North American countries given its superior safety profile. However the high cost of recombinant FVIII, which may be as much as 20-50% higher than plasma-derived FVIII[36], makes it an unaffordable option for a majority of patients around the world. Recent years have seen the attempts to bioengineer variant FVIII with improved characteristics[45] such increased synthesis and expression, improved bioactivity, extended half-life and reduced antigenicity. If successful, future development of improved versions of bioengineered recombinant FVIII with desirable characteristics such as longer half-life and reduced

immunogenicity would certainly be heralded as a step forward for hemophilia treatment and could help to moderate the cost of treatment provided it translates to requiring lower and less frequent doses of FVIII replacement.

1.4.5.3. Synthetic drugs for hemophilia treatment

The discovery in 1977 that desmopressin (1-deamino-8-D-arginine vasopressin, DDAVP) could be used to increase plasma FVIII levels in mild hemophilia and patients with Von Willebrand disease was yet another significant contribution towards hemophilia treatment. Desmopressin, a synthetic analog of the antidiuretic hormone L-arginine vasopressin (AVP), mediates the release of FVIII, VWF and plasminogen activator from storage sites. It therefore serves as an effective but temporary means to elevate plasma FVIII levels 2 to 3 times above basal levels in patients with mild and moderate hemophilia and has been effectively used to halt hemorrhages. However, efficacy is limited by endogenous levels of factor FVIII and the time required to regenerate these factors *in vivo*[42].

Antifibrinolytic drugs[46] are another class of compounds commonly used for control of hemostasis. As the name suggests, these agents prevent or reduce fibrinolysis (the breakdown of fibrin clots). Aprotinin, a bovine-derived protease inhibitor which directly inhibits the fibrinolytic enzyme plasmin, is an FDA-approved hemostatic drug used to reduce bleeding during coronary artery bypass surgery. Epsilon-amino-caproic acid (EACA, Amicar) and tranexamic acid are lysine analogues which bind to fibrin and promote clot stability by inhibiting the activation of a plasminogen, are often used to control bleeds from mucosal surfaces.

Fibrin sealants, mixtures of fibrinogen and thrombin, which are capable of controlling hemostasis by forming fibrin clots when applied to sites of mucosal bleeds, are often used in procedures such as dental extraction and circumcision. Calcium alginate, a polysaccharide extract from brown seaweed, is used as a hemostatic agent in wound dressings or for treating epistaxis, given its ability to control bleeding by exchanging sodium ions at sites of bleed with its calcium ions. Other drugs that can be used to control the severity of bleeds are danazol[47], an attenuated androgen, which increases FVIII levels in mild hemophilia; prednisone, an anti-inflammatory synthetic glucocorticoid that has been successfully used for treating hematuria[48].; and aminoglycosides, such as gentamycin, which increase FVIII levels in patients with nonsense/frameshift FVIII mutations by correcting premature stop codons via alternate usage of amino acids[49].

1.4.5.4. FVIII bypass treatment

In recent years, the use of FVIII bypass agents to restore normal hemostasis has became available and is being actively explored especially in patients with FVIII inhibitors that render conventional FVIII replacement therapy ineffective. Approximately a third of severe hemophilia A patients develop inhibitory antibodies against FVIII and, depending on the severity of inhibiter formation, could benefit from FVIII bypass agents. The rationale of FVIII bypass treatments using Activated Prothrombin Complex (aPCC) or Factor Eight Inhibitor Bypassing Activity (FEIBA) concentrates [50] is to activate the coagulation cascade without the requirement for FVIII activity. These concentrates, consisting of plasma-derived factor II (prothrombin), FVII, FIX and FXa, directly initiate thrombin generation.

Recombinant activated FVII (FVIIa, NovoSeven)[51], which initiates coagulation without the requirement for FVIIIa, has been tested in clinical trials and shown to function as an effective bypass agent for hemophilia. FVIIa activates a key coagulation factor, FX, by binding to TF in subendothelial layers of exposed and injured blood vessels or by directly binding to membranes of activated platelets recruited to sites of vascular injury, thereby leading to thrombin generation and hemostasis. The use of recombinant activated FVII appears to be the best line of treatment for patients with severe hemophilia complicated by high titer inhibitory autoantibodies.

While these bypass agents are generally safe, their use has been associated with thrombosis, pulmonary embolism and cardiovascular complications[46].

1.4.5.5. Gene and cell therapy for hemophilia A

Advances and innovations in medical sciences have significantly improved the quality of life and life expectancy of people with hemophilia around the world. What was once considered a life limiting disease is now regarded as a manageable disorder having a median life expectancy of about 63 years with proper treatment, even in severe hemophilia.[27] . However despite significant progress in hemophilia care, it is estimated that two-thirds of patients in developing countries are either undiagnosed or, if diagnosed, receive no or suboptimal treatment and care[29]. Prohibitively high costs of FVIII replacement therapy combined with unavailability of quality products in developing nations have prompted many researchers to investigate gene and cell therapy approaches as potential curative options. Thus, with the discovery and cloning of FVIII cDNA, in the early 1980s, the battle to develop a cell and gene therapy cure for hemophilia A began to take shape. As of 2013, many techniques for tackling FVIII deficiencies have been tried in laboratory experiments.

To date, a total of three phase I clinical trials have been attempted to treat human subjects[52].

Several features of hemophilia A make it an ideal candidate for gene therapy. The monogenic nature of the disease makes gene augmentation with a functional version of FVIII gene a straightforward and compelling strategy, especially for treating severe hemophilia. Unlike other single protein deficiencies such as insulin, where tight physiological regulation of expression is important, FVIII expression is not finely regulated. Thus, constitutive expression of transgenic FVIII could offer significant phenotypic correction and clinical benefits even with a modest increase in plasma FVIII levels to about 5%,. This alone would improve severe hemophilia to a milder phenotype. A successful gene therapy regime for hemophilia A could be expected to reduce bleeding diathesis and decrease the need for FVIII replacement products. Thus could the high cost of treatment be mitigated and the limited supply of quality FVIII products become of much less concern; these being the two main barriers to good hemophilia care in developing countries. Constitutive FVIII expression by gene therapy will also reduce the requirement for regular venous access for FVIII infusions, an inconvenient, invasive and traumatic experience, especially for children. Several *in vitro* and *in vivo* gene transfer studies have shown that FVIII can be synthesized and secreted as a fully functional protein by several cell types suggesting that ultimately several different organs may be targeted for gene transfer and transgene expression. The existence of various animal models of severe hemophilia A[53] (mice, dogs and sheep) that closely mimic the human disorder provides a sound platform for researchers to test potential therapies and monitor long term safety, thus easing the transition from exploratory research to clinical trials.

1.4.6. Gene therapy options being explored

Apart from liver transplantation, gene therapy to supply deficient clotting factors appears to be the only other feasible approach that could potentially provide a long term cure for hemophilia. Kaufman and colleagues demonstrated for the first time in 1990 *in vitro* secretion of FVIII from cell lines that had been retrovirally transduced to express human FVIII cDNA[54], thus providing experimental evidence to support the feasibility of recombinant FVIII production technologies, and gene and cell therapy for hemophilia. What followed over the decades were the development and testing of various gene therapy strategies to express FVIII either directly in target organs and tissues *in vivo* or in various different cell types *ex vivo* that could ultimately serve as cellular vehicles for FVIII expression when implanted *in vivo*.

Various viral and non-viral methods of gene transfer were evaluated for efficacy and safety. The potential gene therapy options currently being explored for hemophilia treatment will be reviewed in this section.

1.4.6.1. *In vivo* vector delivery

Direct *in vivo* gene transfer usually utilizing a viral vector, or less frequently, a non-viral vector, to transduce or transfect tissues and organs, mainly liver and muscle, to express and secrete transgenic FVIII is one of many options that has been actively explored. Greater success to varying degrees has generally been reported with viral vectors, namely retroviral, lentiviral, adenoviral and adeno-associated viral vectors. To a lesser extent, non-viral methods such as sonoporation, electroporation and hydrodynamic gene delivery of plasmid vectors have also been successfully applied *in vivo*.

One of the most frequently utilized viral gene delivery approaches is systemic administration of MoMLV gamma retroviral vectors. Retroviral vectors effectively transduce a wide variety of cell types, and can theoretically enable durable transgene expression owing to their integration into the genome. Retroviruses encoding human FVIII cDNA have been successfully utilized for durable liver directed expression of transgenic FVIII in animals such as the mice, rabbits and dogs[55]. Based on promising results in animal models, a clinical trial of gamma retroviral vectors for gene transfer of FVIII cDNA was conducted in severe hemophilia A patients[56]. Stable transgene integration was documented in peripheral blood mononuclear cells but the desired therapeutic efficacy was not achieved. Retroviruses are more effective in transducing actively dividing cells and thus *in vivo* delivery and expression strategies using these viruses, in general, have not met with much success. Retroviruses may be more appropriate for *ex vivo* cell transduction, although there are justifiable concerns regarding their potential oncogenicity.

Adenoviruses, which are mainly maintained as episomes and can effectively transduce non-dividing and dividing cells, have also been employed *in vivo* to target FVIII transgene expression mainly in liver and muscle of FVIII-deficient mice and dogs[57]. A major drawback of the early generation adenoviral vectors was the tendency to elicit strong inflammatory responses and hepatotoxicity when delivered *in vivo*[58]. Improved versions of adenoviruses, known as high capacity (HC) adenoviruses, are devoid of most viral genes and shown to be safer[59]. HC adenoviral vectors have been used to express FVIII in murine and canine models of hemophilia A, although the observed therapeutic effects were only transient in some animals owing either to the

development of neutralizing FVIII antibodies or immune-mediated destruction of gene modified cells. A clinical trial of adenoviral vectors for liver directed FVIII expression was prematurely terminated after treating only a single patient with severe hemophilia because of vector related immune response and toxicity[52]. Given the adverse systemic complications reflected by transaminitis (liver toxicity) and thrombocytopenia (bone marrow toxicity) reported with HC adenoviral vectors in at least two other clinical trials for other disorders[58, 60], the risks associated with the use of these vectors must be carefully re-evaluated until further improvements reduce their immunogenicity.

Recombinant adeno-associated viral (rAAV) vectors appear to provide the most promising evidence for efficient and safe *in vivo* transduction and long term transgene expression in both preclinical animal models (rodents, dogs and non-human primates) and human subjects. AAV which is inherently non-pathogenic and non-replicating can transduce a variety of cell types depending on the serotype and generally maintains transgenes as episomes without genomic integration. AAV2[61] and AAV8[62] have been used successfully to express transgenes such as FIX in the muscle and liver, respectively. Although muscle directed AAV-mediated expression tends to elicit a local immune response leading to antibody production correlated with vector dose, this could be circumvented with transient immunosuppression and by intravascular delivery[63] of AAV to transduce a widespread area of muscle to obviate the need for multiple localized intramuscular injection of the virus. Most recently, successful and durable liver directed FIX expression was reported in six hemophilia B patients treated with FIX encoding AAV8 vectors[62]. A caveat to using AAV vectors is the potential for immune response against the capsid proteins which may render the treatment ineffective[64]. For instance, the high prevalence infection by naturally occurring AAV2 and AAV3 in human populations may preclude efficient gene transfer with this AAV serotype due to the preexistence of neutralizing antibodies against the AAV2 capsid. Alternate AAV serotypes such as AAV4 and AAV8 which have limited cross reactivity towards AAV2 serotype may be used to circumvent this problem. Despite promising results, immune clearance of transduced hepatocytes, transaminitis and inflammatory responses are still relevant concerns with this class of vectors[64].

Of the non-viral vector systems and methods being developed and tested for *in vivo* delivery of FVIII, it is worth noting two promising approaches. One is the hydrodynamic intravenous delivery of naked plasmid DNA targeting expression in the liver and muscle; another is the use of integrating non-viral vectors such as

Sleeping Beauty transposons that could potentially mediate long term transgene expression. However, although both approaches have successfully been used in combination to achieve efficacy in hemophilic mice[65], results in larger animal models have been less than satisfactory. A general strategy for improving gene transfer to selective organs such as specific segments of the liver or muscle are being tested in large animal models such as the dogs, non-human primates and pigs. Minimally invasive hydrodynamic gene delivery to liver segments has also been demonstrated in humans[66]. Continuous improvements to *in vivo* gene delivery as well as derivation of superior non-viral integrating vector systems hold promise for development of feasible non-viral gene therapy for disorders such as hemophilia.

1.4.6.2. *Ex vivo* gene-based cellular therapy

While direct *in vivo* transduction of target tissues and cells appears to be an appealing gene augmentation approach for correcting protein deficiencies, these approaches also carry undesirable potential risks such as invoking immune response, unrestricted dissemination of vectors and transduction of unintended organs and cells. In contrast, *ex vivo* gene transfer strategies enable manipulation of specific cell types (depending on disease being treated and protein being expressed) under controlled conditions and, more importantly, allows gene modified cells to be comprehensively characterized and evaluated for genotoxic potential before administration to patients. Unlike the essential irreversibility of direct *in vivo* transduction, treatment plans can be aborted if danger signals are identified *ex vivo*.

Gene delivery, either with viral vectors or by non-viral means such as electroporation or chemical-based transfection, has been successfully performed in a variety of primary human cell types[67-71]. With respect to hemophilia therapy, FVIII cDNA has been delivered and maintained as episomes[72] or integrated into the genome using retroviral[73] and lentiviral vectors,[74] and non-viral vectors integrated by SB-transposons and phiC31 integrase[75].

The future of *ex vivo* cell based therapy for correcting genetic deficiencies appears to be cautiously promising. Demonstration of long term correction of several genetic disorders such as ADA-SCID[76], IL2RG-SCID[77], CGD[78] and WAS[79] lends credibility to the feasibility of *ex vivo* cell therapy. However, issues pertaining to biosafety of some of the approaches need to be carefully evaluated and ironed out to prevent a repeat of adverse events that have plagued several of these clinical trials. The use of viral vectors with reduced immunogenicity and with safer biosafety profiles such as foamy viruses[80] and SIN-viruses[81] may prove to be crucial in

developing safer vectors. ZFNs[82] and TALENs[83] capable of inducing homology-directed integration of transgenes may also be developed as novel non-viral approaches for safe integration and durable expression of transgenes. With much research focus in induced pluripotent stem cells, conventional *ex vivo* cellular therapy approaches utilizing somatic or stem cells may in future also include well characterized iPSCs[84].

The source of cell type for *ex vivo* cell therapy depends on the disease being treated, the type (e.g. the need for post-translational protein modification) and localization of the transgenic protein (e.g. whether expressed intracellularly in specific cells or secreted systemically) and on the availability and ease of isolation, culture and expansion of cells to be used for therapy. The section below summarizes the cell types that have been successfully used in *ex vivo* cell therapy approaches to date, and reviews potential cell types that could be of use in the future.

1.4.6.2.1.   Dermal fibroblasts and epidermal keritinocytes

Primary dermal fibroblasts were an obvious choice for cell therapy as they are easily obtained, cultured and expanded to numbers relevant for therapeutic applications. Early studies also showed that these cells efficiently synthesize, process and secrete transgenic FVIII[70]. Given these favorable characteristics, it is not surprising that fibroblasts were chosen for the first *ex vivo* clinical trial for hemophilia A treatment[85]. However, it was noted that the therapeutic efficacy did not last beyond the first 10 months after treatment, in this clinical study. Apart from reporting FVIII expression in hemophilic mice from direct transduction of epidermal keratinocytes[86], no further progress has been reported with either fibroblasts or keratinocytes for FVIII expression.

1.4.6.2.2.   Hematopoietic stem cells

Gene modified hematopoietic stem cells have been successfully used in *ex vivo* cell therapy for treatment of several immunodeficiency disorders such as ADA and *IL2RG*-SCID, CGD and WAS. Prospects for using HSCs as cellular carriers of transgenes are appealing because gene modified HSCs have the ability to self-renew *in vivo* and thus could persist for a lifetime. Efficient gene transfer can be achieved in HSCs using retroviral or lentiviral vectors and, to a lesser degree, with non-viral methods such as nucleofection of plasmids. HSCs transfected or transduced with FVIII cDNA have been reported to secrete FVIII, albeit at lower levels compared to non-hematopoietic cells[74]. This shortcoming has been partly addressed by engineered

FVIII variants which have superior expression characteristics[87-90]. Long term correction of murine hemophilia by implantation of stably transduced, FVIII-expressing HSCs has been reported[91]. Besides HSCs, differentiated hematopoietic cells such as megakaryocytes/platelets and monocytes have also been evaluated for FVIII expression[92]. While megakaryocytes and monocytes efficiently secreted FVIII, platelets were found to predominantly store FVIII in granules but did partially improve blood clotting times in hemophilic mice even in the presence of high-titer anti-FVIII inhibitory antibodies, presumably due to ectopic expression of FVIII at sites of vascular injury[93]. In summary, hematopoietic stem cell therapy (HSCT) appears to be an attractive option for hemophilia treatment and there are plans to test this approach in human clinical trials[52]. However, a caveat to HSCT is that effective engraftment and *in vivo* reconstitution of HSCs requires pre-conditioning of the patients by radiation or chemotherapeutic agents in order to kill off resident HSCs and induce the process of reconstitution. Oncogenicity arising from insertional mutagenesis associated with retroviral and lentiviral vectors is another prevailing concern with current HSCT protocols.

1.4.6.2.3.   Blood-outgrowth endothelial cells

Blood-outgrowth endothelial cells (BOECs), also known as endothelial progenitor cells (EPCs), belong to a unique family of endothelial marker-expressing, highly proliferative adherent cells derived from long term culture of cord or peripheral blood. While there remain some disagreements with respect to their origin, proliferative potential and differentiation capacity, several groups have demonstrated successful isolation, culture and expansion of BOECs. Estimated to be present in 0.0001% of the mononuclear cell population in peripheral blood, BOECs are collagen adherent cells that exponentially proliferate after an initial lag phase of two weeks in culture. BOECs, initially described by Hebbel and colleagues[94], have been characterized by their expression of endothelial markers such as VE-cadherin, thrombomodulin, VWF, flk-1, PECAM, CD34, P1H12 and CD36, and have excellent proliferative capacity *in vitro*[95]. Several studies have estimated 1.2×10^6-fold [96] to 3.4×10^{12}-fold expansion of these cells in culture which, if correct, means that up to 10^{16} to 10^{19} cells (sufficient for most cell therapy applications) may be produced from x 50 ml of peripheral blood, [94, 97]. Efficient gene transfer (with both viral and non-viral methods) and expression of FVIII has been described for these cells[69, 98]. The efficacy of using BOECs stably secreting FVIII has been successfully demonstrated in hemophilic mice for up to 6 months[97]. Although BOECs have several appealing

characteristics that warrant their serious consideration for cell therapy applications, a search of the literature indicates that the feasibility of translating such an approach to a larger animal model has not yet been attempted with equal success.

1.4.6.2.4. Bone marrow stromal cells

Bone marrow stromal cells (BMSCs) are another unique group of multipotent, clonogenic, adherent, fibroblast-like cells that have potential for multiple therapeutic applications. BMSCs can readily be harvested by simple non-invasive aspiration of bone marrow, established in culture and expanded extensively through multiple population doublings *in vitro* to derive about 0.5×10^8 cells in less than a month[99]. Efficient gene transfer to BMSCs by non-viral means such as nucleofection[100] or with retro- and lentiviral vectors[73] has been extensively reported as has the potential for gene modified BMSCs to efficiently secrete transgenic FVIII. Several studies have demonstrated the potential for durable FVIII secretion and long term phenotypic correction in hemophilic mice[101] and sheep[102] implanted with autologous gene modified BMSCs. Given the ease of isolation and culture, capacity for deriving sufficient cell numbers, efficient gene transfer, and propensity for durable FVIII secretion coupled with demonstration of efficacy in preclinical animal models, BMSCs could be ideal cellular vehicles for FVIII replacement therapy in humans. *Ex vivo* cultured BMSCs have safely been used in human subjects for regenerative therapy of bone, cartilage and muscle defects[103]. In summary, MSCs have a broad range of potential applications in gene and cell therapy of various genetic and degenerative medical conditions. Impressive results from studies of preclinical animal models of hemophilia lend credence to attempts to develop similar therapies for human subjects. However, there are unresolved and controversial issues such as heterogeneity between donor samples, amenability for *ex vivo* expansion and genomic stability of long term cultured cells[104] (malignant transformation subsequent to clonal selection of cells extensively cultured *ex vivo*) to be carefully addressed before MSCs find widespread therapeutic uses in human subjects.

1.4.6.2.5. Cord-lining epithelial cells

Neonatal cells derived from the umbilical cord are another source of potentially useful cell types that could be developed for autologous cell therapy, especially for pediatric patients with inborn genetic disorders. Umbilical cord blood and placenta have long been rich sources of stem cells such as CD34+, mesenchymal, epithelial and endothelial progenitor cells, to name just a few. Stem cells isolated,

expanded and cryopreserved from umbilical cords of newborn infants may serve as useful reserves of autologous cells to be used for future gene and cell therapy, if the need arises. The challenge with umbilical cord derived cells, as with all other adult cells, is the need for *ex vivo* expansion of these cells to numbers that are realistic for therapeutic applications. This is estimated to be in the range of 1 to 10 x10^7 cells, and even more if repeated treatment is envisaged. Dr. Phan Toan Tang and colleagues have described the isolation and characterization of a novel epithelial-like cell type derived from the outer lining membrane of human umbilical cords[105, 106]. Cord-lining epithelial cells (CLECs) have several favorable characteristics that make them putative candidates for autologous cell therapy. CLECs grow as anchorage dependent cells *in vitro*, express pluripotency markers such as *Oct-4* and *Nanog* and have excellent proliferative capacity. Approximately 6 x 10^9 fresh CLECs can be harvested from a single umbilical cord and these can be expanded for another 30 to 40 population doublings to yield cell numbers that are more than sufficient for therapeutic applications. The use of umbilical cord-derived cells also does not attract the same ethical and safety issues as embryonic stem cells. CLECs can be efficiently transfected by non-viral means and my work has shown them to be capable of synthesizing, processing and secreting FVIII at therapeutically meaningful levels when transfected with FVIII cDNA[75]. Immunological characterization of CLECs showed that these cells express immunomodulatory proteins such as HLA-E and HLA-G. Furthermore, CLECs were demonstrated to inhibit T lymphocyte response in a mixed leukocyte reaction assay[105] suggesting their potential also for allogeneic cell therapy. To date, CLECs have been investigated for various potential applications such as epidermal reconstitution[107], ocular surface regeneration for treating ocular disorders[106] and as bio-implants for treating metabolic and genetic disorders[75, 105]. Banking of cord-lining stem cells for future application are now available in several countries such as Vietnam (MekoStem), Taiwan (Stemcyte) and Singapore (Cordlife).

1.4.7. Bioengineered superior variants of FVIII

One of the challenges with transgenic FVIII expression *in vitro* and *in vivo* is the difficulty in obtaining sufficiently high levels of the transgenic protein. Transgenic FVIII expression levels are 2 to 3 orders of magnitude lower than other proteins of similar size. Biochemical features inherent in FVIII contribute to inefficient transcription, poor translation, folding and secretion of the protein[108]. Cis-elements which act as transcriptional repressors[109] and dominant inhibitors of RNA accumulation[110] contributing to inefficient mRNA transcription have been reported in

FVIII cDNA. Misfolding and premature degradation of primary translation products are hypothesized as reasons for low expression levels [111]. FVIII secretion is also thought to be inefficient due to retention in the endoplasmic reticulum (ER) through interaction with ER chaperone proteins such as immunoglobulin-binding protein (BiP), calnexin and calreticulin[108]. Furthermore, transit of FVIII from ER to the Golgi apparatus, which is speculated to be the rate-limiting step in secretion, requires a facilitated transport mechanism *via* interaction with mannose-binding lectin 1 (LMAN1)[22]. These factors collectively contribute to an inefficiently synthesized and secreted FVIII protein, a shortcoming that contributes to high cost of recombinant FVIII production and which undermines gene therapy approaches that require FVIII expression. Another limitation of the FVIII protein is its shortlived activity, a consequence of proteolytic inactivation that follows immediately after its activation by thrombin. Having identified the molecular mechanisms contributing to low FVIII expression, several investigators attempted to bioengineer FVIII variants with enhanced characteristics such as improved transcription and secretion, increased resistance to inactivation and extended plasma half-life. Another favorable characteristic worth modifying is to reduce the antigenicity and immunogenicity of FVIII protein.

The first improved FVIII variant reported was derived by deleting the entire B-domain which is encoded by 38% of full length FVIII cDNA. This not only reduced the FVIII cDNA size to one that facilitated cloning into viral vectors with limited cargo capacity but also increased both the mRNA and secreted FVIII protein levels by 20-fold and 3-fold, respectively[112]. A single mutation that converts phenylalanine at position 309 in the A1 domain of FVIII to serine (F309S) or alanine (F309A) significantly improved FVIII secretion by approximately 3-fold. Transit of FVIII from ER to Golgi and its subsequent secretion is limited by its interaction with BiP. This is exacerbated by the requirement of high intracellular levels of ATP for its release. The F309S mutation which resides within the BiP-interacting domain was shown to reduce the ATP requirement for release of FVIII protein from BiP, and thus improve its secretion[88].

Previous studies have shown that LMAN1 facilitates transport of FVIII from ER to Golgi *via* interaction with N-linked oligosaccharides (glycosylation sites) in the B domain. Miao and colleagues showed that retaining a minimum of 6 to 8 N-linked glycosylation sites within the B domain (approximately 33% of the B domain) could improve FVIII secretion by 4- to 9-fold compared to a FVIII with a completely deleted B domain[90].

Improved FVIII expression has also been reported after codon optimization of human B domain-deleted (BDD)-FVIII cDNA[113] and with the use of BDD-hybrid FVIII comprising human and porcine FVIII [114]. The former modification improved FVIII expression by up to 44-fold compared with unmodified human FVIII, while the latter modification of replacing human A1 and A3 domains with homologous porcine domains resulted in 10- to 100-fold increased FVIII expression[114, 115] presumably due to more efficient post-translational transit of the hybrid FVIII protein through the secretory pathway[116].

FVIII variants with enhanced stability and resistance to inactivation have been derived by replacing charged residues in the A2 domain with hydrophobic ones at the interface of the A1 and A3 domains[117]. Inactivation-resistant FVIII, known as IR8, with A2 domain covalently linked to the light chain (A3-C1-C2) was derived by partial amino acid deletions (794-1689) that create missense mutations at cleavage sites for thrombin and activated protein C. IR8 was shown to retain FVIII activity significantly longer *in vitro* and retained the ability to bind to VWF[111]. FVIII engineered with disulfide bonds between A2 and A3 domains (C664 and C1826) also had more stable FVIII activity due to slower dissociation of the A2 domain, and thus slower proteolysis and inactivation[118].

Attempts have also been made to engineer variants with longer plasma half-life. FVIII is cleared from circulation by binding to hepatic endocytic receptors such LRP and HSPG. FVIII variants modified to weaken or abolish LRP (A2 domain amino acid residues 484-509 and C2 domain) or HSPG (A2 domain amino acid residues 558-565) have improved plasma half-lives. Similar effects were obtained with anti-C2 domain monoclonal antibodies such as ESH4 which inhibit LRP binding to FVIII[17]. Another strategy for improving plasma half-life involves conjugation of FVIII to PEG polymers. PEGylation blocks interaction with clearance receptors and reduces formation of neutralizing antibodies[119].

Continuous efforts are being invested in developing and testing FVIII with reduced immunogenicity and antigenicity in order to reduce the risk of inhibitory autoantibodies against FVIII. For instance, recombinant BDD-porcine FVIII (OBI-1) is being evaluated in a clinical trial of patients with inhibitory antibodies[120] . Variants with reduced antigenicity are being developed by replacing key residues within the A2 domain that are common epitopes for inhibitory antibodies. R484A/R489A/P492A mutant FVIII with mutations in the A2 domain were less immunogenic in hemophilic mice compared to BDD-FVIII[121].

1.4.8. Gene and cell therapy clinical trials for hemophilia A

Encouraged by durable expression of high levels of transgenic FVIII *in vitro* in several primary human cell types as well as *in vivo* in murine and canine hemophilic animal models, three independent corporate-sponsored phase I human clinical trials for hemophilia A were initiated with much enthusiasm.

The first trial, initiated by Transkaryotic Therapy Inc., involved implantation of six patients with severe hemophilia with transgenic FVIII-secreting autologous dermal fibroblasts that had been electroporated with a BDD-human FVIII cDNA. Clonal cells with stable integration of the transgene and demonstrating high levels of FVIII secretion were expanded and surgically implanted into the omentum. The trial reported a modest but significant increase in plasma FVIII levels (0.5 – 4% of normal) in four of 6 patients and concomitant reduced bleeding episodes and decreased requirements for FVIII replacement for up to 10 months after treatment[85]. Although the reasons for lack of therapeutic efficacy beyond 10 months were not determined, turnover and eventual death of implanted cells as well as promoter/transgene silencing could be speculated as possible reasons.

The second clinical study, a phase I clinical trial conducted by Chiron Inc., employed intravenous delivery of MoMLV retroviral vectors encoding BDD-human FVIII in 13 adults with severe hemophilia. Peripheral blood mononuclear cells showed evidence of vector DNA up to 1 year post-treatment. However, apart from the rare and occasional transient rise in plasma FVIII levels, potentially fewer bleeding episodes and slightly diminished requirements for FVIII replacement, no sustained FVIII expression was observed in all the treated patients during the course of the study[56].

The third and last hemophilia A trial by GenStar Therapeutics Inc., involved intravenous delivery of a HC adenoviral vector encoding full length human FVIII cDNA with the intent of targeting the liver for transgene expression. Plasma FVIII levels in the range of 1% were detected for several months in the single and only patient recruited for this trial[52]. However, the trial was halted indefinitely because of vector related adverse events (fever, transient hepatotoxicity, thrombocytopenia) that were probably consequences of immune responses evoked by the adenoviral vector.

Lessons learned from these early gene therapy attempts to treat hemophilia were the need to develop strategies that minimize the risk of eliciting a strong immune response and which achieve higher and sustained expression of FVIII protein at therapeutically relevant levels. Viral vectors with improved safety profiles capable of transducing a wide range of cell types are being developed and tested. Strategies

involving implantation of *ex vivo* transduced primary cell types including induced pluripotent stem cells and embryonic stem cells are also actively being explored in preclinical animal models for long term efficacy. Novel FVIII molecules bioengineered for lower immunogenicity, extended half- life and improved secretion or activity are being investigated.

Despite the disappointing results from the aforementioned clinical trials, novel and improved gene and cell therapy approaches for hemophilia treatment are continually being developed and tested, with some poised for preclinical and clinical studies. Very recently, encouraging results were reported in a clinical trial for hemophilia B. Patients with severe hemophilia B treated with AAV8 vector encoding human FIX expressed sustained FIX levels between 2% to 11% for up to 16 months[62], with only minor side effects, if any. While such optimistic results could rejuvenate interest in developing more gene therapy clinical trials for hemophilia, exercising caution on behalf of safety would be prudent. Given that hemophilia is a disease that can be treated with conventional protein replacement therapy, the risks of gene and cell therapy approaches must be carefully weighed against the benefits they bring. Bio-safety and efficacy of candidate approaches must be carefully and objectively evaluated in preclinical models before advancing to human trials.

1.4.9. References

1. Mackman N, Tilley RE, Key NS. Role of the extrinsic pathway of blood coagulation in hemostasis and thrombosis. Arterioscler Thromb Vasc Biol. (2007) 27:1687-1693.

2. Gailani D, Renné T. Intrinsic pathway of coagulation and arterial thrombosis. Arterioscler Thromb Vasc Biol. (2007) 27:2507-2513.

3. Jockenhoevel S, Flanagan TC. Cardiovascular Tissue Engineering Based on Fibrin-Gel-Scaffolds, Tissue Engineering for Tissue and Organ Regeneration. In: Prof. Daniel Eberli, editor. Tissue Engineering for Tissue and Organ Regeneration. InTech; 2011.

4. Salomon O, Zivelin A, Livnat T, Seligsohn U, *et al*. Inhibitors to Factor XI in patients with severe Factor XI deficiency. Semin Hematol. (2006) 43:S10-S12.

5. Lämmle B, Wuillemin WA, Huber I, Krauskopf M, *et al*. Thromboembolism and bleeding tendency in congenital factor XII deficiency--a study on 74 subjects from 14 Swiss families. Thromb Haemost. (1991) 65:117-121.

6. Stonebraker JS, Bolton-Maggs PHB, Soucie JM, Walker I, *et al*. A study of variations in the reported haemophilia A prevalence around the world. Haemophilia. (2010) 16:20-32.

7. Mannucci PM, Tuddenham EG. The hemophilias--from royal genes to gene therapy. N Engl J Med. (2001) 344:1773-1779.

8. Sborov DW, Rodgers GM. Acquired hemophilia a: a current review of autoantibody disease. Clin Adv Hematol Oncol. (2012) 10:19-27.

9. Franchini M, Mannucci PM. Past, present and future of hemophilia: a narrative review. Orphanet J Rare Dis. (2012) 7:24.

10. Patek AJ, Taylor FH. Hemophilia. II. Some properties of a substance obtained from normal human plasma effective in accelerating the coagulation of hemophilic blood. J Clin Invest. (1937) 16:113-124.

11. Pool JG, Gershgold EJ, Pappenhagen AR. High-potency antihaemophilic factor concentrate prepared from cryoglobulin precipitate. Nature. (1964) 203:312.

12. Lillicrap D. Extending half-life in coagulation factors: where do we stand? Thromb Res. (2008) 122 Suppl 4:S2-S8.

13. Hoyer LW. Hemophilia A. N Engl J Med. (1994) 330:38-47.

14. Vehar GA, Keyt B, Eaton D, Rodriguez H*, et al*. Structure of human factor VIII. Nature. (1984) 312:337-342.

15. Toole JJ, Knopf JL, Wozney JM, Sultzman LA*, et al*. Molecular cloning of a cDNA encoding human antihaemophilic factor. Nature. (1984) 312:342-347.

16. Lenting PJ, van Mourik JA, Mertens K. The life cycle of coagulation factor VIII in view of its structure and function. Blood. (1998) 92:3983-3996.

17. Lenting PJ, Neels JG, van den Berg BM, Clijsters PP*, et al*. The light chain of factor VIII comprises a binding site for low density lipoprotein receptor-related protein. J Biol Chem. (1999) 274:23734-23739.

18. Sarafanov AG, Ananyeva NM, Shima M, Saenko EL*, et al*. Cell surface heparan sulfate proteoglycans participate in factor VIII catabolism mediated by low density lipoprotein receptor-related protein. J Biol Chem. (2001) 276:11970-11979.

19. Bovenschen N, Herz J, Grimbergen JM, Lenting PJ*, et al*. Elevated plasma factor VIII in a mouse model of low-density lipoprotein receptor-related protein deficiency. Blood. (2003) 101:3933-3939.

20. Oldenburg J, Ananyeva NM, Saenko EL. Molecular basis of haemophilia A. Haemophilia. (2004) 10 Suppl 4:133-139.

21. Mazurier C, Dieval J, Jorieux S, Delobel J*, et al*. A new von Willebrand factor (vWF) defect in a patient with factor VIII (FVIII) deficiency but with normal levels and multimeric patterns of both plasma and platelet vWF. Characterization of abnormal vWF/FVIII interaction. Blood. (1990) 75:20-26.

22. Nichols WC, Seligsohn U, Zivelin A, Terry VH, *et al*. Mutations in the ER-Golgi intermediate compartment protein ERGIC-53 cause combined deficiency of coagulation factors V and VIII. Cell. (1998) 93:61-70.

23. Mannucci PM. Back to the future: a recent history of haemophilia treatment. Haemophilia. (2008) 14 Suppl 3:10-18.

24. Jones P, Robillard L. The World Federation of Hemophilia: 40 years of improving haemophilia care worldwide. Haemophilia. (2003) 9:663-669.

25. Azzi A, Morfini M, Mannucci PM. The transfusion-associated transmission of parvovirus B19. Transfus Med Rev. (1999) 13:194-204.

26. Zaman SMA, Hill FGH, Palmer B, Millar CM, *et al*. The risk of variant Creutzfeldt-Jakob disease among UK patients with bleeding disorders, known to have received potentially contaminated plasma products. Haemophilia. (2011) 17:931-937.

27. Larsson SA. Life expectancy of Swedish haemophiliacs, 1831-1980. Br J Haematol. (1985) 59:593-602.

28. Pratt KP. Inhibitory antibodies in hemophilia A. Curr Opin Hematol. (2012) 19:399-405.

29. O'Mahony B, Black C. Expanding hemophilia care in developing countries. Semin Thromb Hemost. (2005) 31:561-568.

30. Coppola A, Tagliaferri A, Di Capua M, Franchini M, *et al*. Prophylaxis in children with hemophilia: evidence-based achievements, old and new challenges. Semin Thromb Hemost. (2012) 38:79-94.

31. Shanbrom E, Thelin GM. Experimental prophylaxis of severe hemophilia with a factor VIII concentrate. JAMA. (1969) 208:1853-1856.

32. Collins PW, Björkman S, Fischer K, Blanchette V, *et al*. Factor VIII requirement to maintain a target plasma level in the prophylactic treatment of severe hemophilia

A: influences of variance in pharmacokinetics and treatment regimens. J Thromb Haemost. (2010) 8:269-275.

33. Medical and Scientific Advisory Council (MASAC) recommendation #179: MASAC recommendation concerning prophylaxis (regular administration of clotting factor concentrate to prevent bleeding). 2007. (Accessed November, 2012).

34. World Health Organization. Delivery of Treatment for Haemophilia: Report of a Joint WHO/WFH/ISTH Meeting London, United Kingdom, 11- 13 February 2002. In. London, United Kingdom: WHO; 2002.

35. Manco-Johnson MJ, Abshire TC, Shapiro AD, Riske B, *et al*. Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. N Engl J Med. (2007) 357:535-544.

36. Mannucci PM. Hemophilia: treatment options in the twenty-first century. J Thromb Haemost. (2003) 1:1349-1355.

37. Tabor E. The epidemiology of virus transmission by plasma derivatives: clinical studies verifying the lack of transmission of hepatitis B and C viruses and HIV type 1. Transfusion. (1999) 39:1160-1168.

38. Mannucci PM. Plasma-derived versus recombinant factor VIII concentrates for the treatment of haemophilia A: plasma-derived is better. Blood transfusion = Trasfusione del sangue. (2010) 8:288-291.

39. Gitschier J, Wood WI, Goralka TM, Wion KL, *et al*. Characterization of the human factor VIII gene. Nature. (1984) 312:326-330.

40. Wood WI, Capon DJ, Simonsen CC, Eaton DL, *et al*. Expression of active human factor VIII from recombinant DNA clones. Nature. (1984) 312:330-337.

41. White GC, McMillan CW, Kingdon HS, Shoemaker CB, *et al*. Use of recombinant antihemophilic factor in the treatment of two patients with classic hemophilia. N Engl J Med. (1989) 320:166-170.

42. Gringeri A, Muça-Perja M, Mangiafico L, von Mackensen S, *et al*. Pharmacotherapy of haemophilia A. Expert Opin Biol Ther. (2011) 11:1039-1053.

43. Franchini M, Lippi G. Von Willebrand factor-containing factor VIII concentrates and inhibitors in haemophilia A. A critical literature review. Thromb Haemost. (2010) 104:931-940.

44. Dasgupta S, Repessé Y, Bayry J, Navarrete AM, *et al*. VWF protects FVIII from endocytosis by dendritic cells and subsequent presentation to immune effectors. Blood. (2007) 109:610-612.

45. Pipe SW. The promise and challenges of bioengineered recombinant clotting factors. J Thromb Haemost. (2005) 3:1692-1701.

46. Mannucci PM, Levi M. Prevention and treatment of major blood loss. N Engl J Med. (2007) 356:2301-2311.

47. Gralnick HR, Rick ME. Danazol increases factor VIII and factor IX in classic hemophilia and Christmas disease. N Engl J Med. (1983) 308:1393-1395.

48. Abildgaard CF, Simone JV, Schulman I. Steroid treatment of hemophilic hematuria. J Pediatr. (1965) 66:117-119.

49. Mammen Chandy. Treatment options in the management of hemophilia in developing countries. World Federation of Hemophilia. Treatment of hemophilia. December 2005 (No. 37).

50. Sjamsoedin LJ, Heijnen L, Mauser-Bunschoten EP, van Geijlswijk JL, *et al*. The effect of activated prothrombin-complex concentrate (FEIBA) on joint and muscle bleeding in patients with hemophilia A and antibodies to factor VIII. A double-blind clinical trial. N Engl J Med. (1981) 305:717-721.

51. Hedner U, Lee CA. First 20 years with recombinant FVIIa (NovoSeven). Haemophilia. (2011) 17:e172-e182.

52. Doering CB, Spencer HT. Advancements in gene transfer-based therapy for hemophilia A. Expert review of hematology. (2009) 2:673-683.

53. Sabatino DE, Nichols TC, Merricks E, Bellinger DA*, et al*. Animal models of hemophilia. Progress in molecular biology and translational science. (2012) 105:151-209.

54. Israel DI, Kaufman RJ. Retroviral-mediated transfer and amplification of a functional human factor VIII gene. Blood. (1990) 75:1074-1080.

55. Van Damme A, Chuah MKL, Collen D, VandenDriessche T*, et al*. Onco-retroviral and lentiviral vector-based gene therapy for hemophilia: preclinical studies. Semin Thromb Hemost. (2004) 30:185-195.

56. Powell JS, Ragni MV, White GC, Lusher JM*, et al*. Phase 1 trial of FVIII gene transfer for severe hemophilia A using a retroviral construct administered by peripheral intravenous infusion. Blood. (2003) 102:2038-2045.

57. Chuah MKL, Schiedner G, Thorrez L, Brown B*, et al*. Therapeutic factor VIII levels and negligible toxicity in mouse and dog models of hemophilia A following gene therapy with high-capacity adenoviral vectors. Blood. (2003) 101:1734-1743.

58. Chuah MKL, Collen D, VandenDriessche T. Biosafety of adenoviral vectors. Curr Gene Ther. (2003) 3:527-543.

59. Segura M, Alba R, Bosch A, Chillon M*, et al*. Advances in helper-dependent adenoviral vector research. Curr Gene Ther. (2008) 8:222-235.

60. Raper SE, Chirmule N, Lee FS, Wivel NA*, et al*. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab. (2003) 80:148-158.

61. Kay MA, Manno CS, Ragni MV, Larson PJ*, et al*. Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. Nat Genet. (2000) 24:257-261.

62. Nathwani AC, Tuddenham EGD, Rangarajan S, Rosales C*, et al*. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med. (2011) 365:2357-2365.

63. Arruda VR, Stedman HH, Haurigot V, Buchlis G, *et al*. Peripheral transvenular delivery of adeno-associated viral vectors to skeletal muscle as a novel therapy for hemophilia B. Blood. (2010) 115:4678-4688.

64. Tenenbaum L, Lehtonen E, Monahan PE. Evaluation of risks related to the use of adeno-associated virus-based vectors. Curr Gene Ther. (2003) 3:545-565.

65. Yant SR, Meuse L, Chiu W, Ivics Z, *et al*. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet. (2000) 25:35-41.

66. Khorsandi SE, Bachellier P, Weber JC, Greget M, *et al*. Minimally invasive and selective hydrodynamic gene therapy of liver segments in the pig and human. Cancer Gene Ther. (2008) 15:225-230.

67. Xu D, Alipio Z, Fink LM, Adcock DM, *et al*. Phenotypic correction of murine hemophilia A using an iPS cell-based therapy. Proc Natl Acad Sci U S A. (2009) 106:808-813.

68. Chuah MK, Van Damme A, Zwinnen H, Goovaerts I, *et al*. Long-term persistence of human bone marrow stromal cells transduced with factor VIII-retroviral vectors and transient production of therapeutic levels of human factor VIII in nonmyeloablated immunodeficient mice. Hum Gene Ther. (2000) 11:729-738.

69. van den Biggelaar M, Bouwens EAM, Kootstra NA, Hebbel RP, *et al*. Storage and regulated secretion of factor VIII in blood outgrowth endothelial cells. Haematologica. (2009) 94:670-678.

70. Zatloukal K, Cotten M, Berger M, Schmidt W, *et al*. In vivo production of human factor VII in mice after intrasplenic implantation of primary fibroblasts transfected by receptor-mediated, adenovirus-augmented gene delivery. Proc Natl Acad Sci U S A. (1994) 91:5148-5152.

71. Kasuda S, Kubo A, Sakurai Y, Irion S, *et al*. Establishment of embryonic stem cells secreting human factor VIII for cell-based treatment of hemophilia A. J Thromb Haemost. (2008) 6:1352-1359.

72. Kurosaki H, Hiratsuka M, Imaoka N, Iida Y, *et al*. Integration-free and stable expression of FVIII using a human artificial chromosome. J Hum Genet. (2011) 56:727-733.

73. Chuah MK, Brems H, Vanslembrouck V, Collen D, *et al*. Bone marrow stromal cells as targets for gene therapy of hemophilia A. Hum Gene Ther. (1998) 9:353-365.

74. Tiede A, Eder M, von Depka M, Battmer K, *et al*. Recombinant factor VIII expression in hematopoietic cells following lentiviral transduction. Gene Ther. (2003) 10:1917-1925.

75. Sivalingam J, Krishnan S, Ng WH, Lee SS, *et al*. Biosafety assessment of site-directed transgene integration in human umbilical cord-lining cells. Mol Ther. (2010) 18:1346-1356.

76. Aiuti A, Slavin S, Aker M, Ficara F, *et al*. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science. (2002) 296:2410-2413.

77. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, *et al*. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest. (2008) 118:3132-3142.

78. Ott MG, Schmidt M, Schwarzwaelder K, Stein S, *et al*. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med. (2006) 12:401-409.

79. Boztug K, Schmidt M, Schwarzer A, Banerjee PP, *et al*. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N Engl J Med. (2010) 363:1918-1927.

80. Erlwein O, McClure MO. Progress and prospects: foamy virus vectors enter a new age. Gene Ther. (2010) 17:1423-1429.

81. Avedillo Díez I, Zychlinski D, Coci EG, Galla M, *et al*. Development of novel efficient SIN vectors with improved safety features for Wiskott-Aldrich syndrome stem cell based gene therapy. Mol Pharm. (2011) 8:1525-1537.

82. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, *et al*. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. (2010) 11:636-646.

83. Mussolino C, Morbitzer R, Lütge F, Dannemann N, *et al*. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. (2011) 39:9283-9293.

84. Liras A. Induced human pluripotent stem cells and advanced therapies: future perspectives for the treatment of haemophilia? Thromb Res. (2011) 128:8-13.

85. Roth DA, Tawa NE, O'Brien JM, Treco DA, *et al*. Nonviral transfer of the gene encoding coagulation factor VIII in patients with severe hemophilia A. N Engl J Med. (2001) 344:1735-1742.

86. Fakharzadeh SS, Zhang Y, Sarkar R, Kazazian HH, *et al*. Correction of the coagulation defect in hemophilia A mice through factor VIII expression in skin. Blood. (2000) 95:2799-2805.

87. Plantier JL, Rodriguez MH, Enjolras N, Attali O, *et al*. A factor VIII minigene comprising the truncated intron I of factor IX highly improves the in vitro production of factor VIII. Thromb Haemost. (2001) 86:596-603.

88. Swaroop M, Moussalli M, Pipe SW, Kaufman RJ, *et al*. Mutagenesis of a potential immunoglobulin-binding protein-binding site enhances secretion of coagulation factor VIII. J Biol Chem. (1997) 272:24121-24124.

89. Doering CB, Healey JF, Parker ET, Barrow RT, *et al*. High level expression of recombinant porcine coagulation factor VIII. J Biol Chem. (2002) 277:38345-38349.

90. Miao HZ, Sirachainan N, Palmer L, Kucab P, *et al*. Bioengineering of coagulation factor VIII for improved secretion. Blood. (2004) 103:3412-3419.

91. Moayeri M, Hawley TS, Hawley RG. Correction of murine hemophilia A by hematopoietic stem cell gene therapy. Mol Ther. (2005) 12:1034-1042.

92. Sadelain M, Chang A, Lisowski L. Supplying clotting factors from hematopoietic stem cell-derived erythroid and megakaryocytic lineage cells. Mol Ther. (2009) 17:1994-1999.

93. Shi Q, Wilcox DA, Fahs SA, Weiler H, *et al*. Factor VIII ectopically targeted to platelets is therapeutic in hemophilia A with high-titer inhibitory antibodies. J Clin Invest. (2006) 116:1974-1982.

94. Lin Y, Weisdorf DJ, Solovey A, Hebbel RP, *et al*. Origins of circulating endothelial cells and endothelial outgrowth from blood. J Clin Invest. (2000) 105:71-77.

95. Lin Y, Chang L, Solovey A, Healey JF, *et al*. Use of blood outgrowth endothelial cells for gene therapy for hemophilia A. Blood. (2002) 99:457-462.

96. Stockschlaeder M, Shardakova O, Weber K, Stoldt VR, *et al*. Highly efficient lentiviral transduction of phenotypically and genotypically characterized endothelial progenitor cells from adult peripheral blood. Blood Coagul Fibrinolysis. (2010) 21:464-473.

97. Matsui H. Endothelial progenitor cell-based therapy for hemophilia A. Int J Hematol. (2012) 95:119-124.

98. Matsui H, Shibata M, Brown B, Labelle A, *et al*. Ex vivo gene therapy for hemophilia A that enhances safe delivery and sustained in vivo factor VIII expression from lentivirally engineered endothelial progenitors. Stem Cells. (2007) 25:2660-2669.

99. Sensebé L, Bourin P, Tarte K. Good manufacturing practices production of mesenchymal stem/stromal cells. Hum Gene Ther. (2011) 22:19-26.

100. Aluigi M, Fogli M, Curti A, Isidori A, *et al*. Nucleofection is an efficient nonviral transfection technique for human bone marrow-derived mesenchymal stem cells. Stem Cells. (2006) 24:454-461.

101. Gangadharan B, Parker ET, Ide LM, Spencer HT*, et al*. High-level expression of porcine factor VIII from genetically modified bone marrow-derived stem cells. Blood. (2006) 107:3859-3864.

102. Porada CD, Sanada C, Kuo CJ, Colletti E*, et al*. Phenotypic correction of hemophilia A in sheep by postnatal intraperitoneal transplantation of FVIII-expressing MSC. Exp Hematol. (2011) 39:1124-1135.e4.

103. Sensebé L, Krampera M, Schrezenmeier H, Bourin P*, et al*. Mesenchymal stem cells for clinical application. Vox Sang. (2010) 98:93-107.

104. Bernardo ME, Cometa AM, Pagliara D, Vinti L*, et al*. Ex vivo expansion of mesenchymal stromal cells. Best Pract Res Clin Haematol. (2011) 24:73-81.

105. Zhou Y, Gan SU, Lin G, Lim YT*, et al*. Characterization of Human Umbilical Cord Lining Derived Epithelial Cells and Transplantation Potential. Cell Transplant. (2011)

106. Reza HM, Ng BY, Gimeno FL, Phan TT*, et al*. Umbilical cord lining stem cells as a novel and promising source for ocular surface regeneration. Stem Cell Rev. (2011) 7:935-947.

107. Huang L, Wong YP, Gu H, Cai YJ*, et al*. Stem cell-like properties of human umbilical cord lining epithelial cells and the potential for epidermal reconstitution. Cytotherapy. (2011) 13:145-155.

108. Kaufman RJ, Pipe SW, Tagliavacca L, Swaroop M*, et al*. Biosynthesis, assembly and secretion of coagulation factor VIII. Blood Coagul Fibrinolysis. (1997) 8 Suppl 2:S3-14.

109. Hoeben RC, Fallaux FJ, Cramer SJ, van den Wollenberg DJ*, et al*. Expression of the blood-clotting factor-VIII cDNA is repressed by a transcriptional silencer located in its coding region. Blood. (1995) 85:2447-2454.

110. Lynch CM, Israel DI, Kaufman RJ, Miller AD*, et al*. Sequences in the coding region of clotting factor VIII act as dominant inhibitors of RNA accumulation and protein production. Hum Gene Ther. (1993) 4:259-272.

111. Pipe SW, Kaufman RJ. Characterization of a genetically engineered inactivation-resistant coagulation factor VIIIa. Proc Natl Acad Sci U S A. (1997) 94:11851-11856.

112. Toole JJ, Pittman DD, Orr EC, Murtha P, *et al*. A large region (approximately equal to 95 kDa) of human factor VIII is dispensable for in vitro procoagulant activity. Proc Natl Acad Sci U S A. (1986) 83:5939-5942.

113. Ward NJ, Buckley SMK, Waddington SN, Vandendriessche T, *et al*. Codon optimization of human factor VIII cDNAs leads to high-level expression. Blood. (2011) 117:798-807.

114. Doering CB, Healey JF, Parker ET, Barrow RT, *et al*. Identification of porcine coagulation factor VIII domains responsible for high level expression via enhanced secretion. J Biol Chem. (2004) 279:6546-6552.

115. Doering CB, Denning G, Dooriss K, Gangadharan B, *et al*. Directed engineering of a high-expression chimeric transgene as a strategy for gene therapy of hemophilia A. Mol Ther. (2009) 17:1145-1154.

116. Brown HC, Gangadharan B, Doering CB. Enhanced biosynthesis of coagulation factor VIII through diminished engagement of the unfolded protein response. J Biol Chem. (2011) 286:24451-24457.

117. Wakabayashi H, Varfaj F, Deangelis J, Fay PJ, *et al*. Generation of enhanced stability factor VIII variants by replacement of charged residues at the A2 domain interface. Blood. (2008) 112:2761-2769.

118. Gale AJ, Pellequer JL. An engineered interdomain disulfide bond stabilizes human blood coagulation factor VIIIa. J Thromb Haemost. (2003) 1:1966-1971.

119. Mei B, Pan C, Jiang H, Tjandra H, *et al*. Rational design of a fully active, long-acting PEGylated factor VIII for hemophilia A treatment. Blood. (2010) 116:270-279.

120. Toschi V. OBI-1, porcine recombinant Factor VIII for the potential treatment of patients with congenital hemophilia A and alloantibodies against human Factor VIII. Curr Opin Mol Ther. (2010) 12:617-625.

121. Parker ET, Healey JF, Barrow RT, Craddock HN, *et al*. Reduction of the inhibitory antibody response to human factor VIII in hemophilia A mice by mutagenesis of the A2 domain B-cell epitope. Blood. (2004) 104:704-710.

1.5.   Rationale, objectives and scope of project

Durable restoration of FVIII secretion *in vivo* holds much potential to correct FVIII deficiencies and will bring about significant clinical benefits to patients with hemophilia A. Implantation of autologous cells stably modified to express and secrete FVIII is the goal of several research groups. Durable and clinically meaningful reconstitution of FVIII expression with negligible genotoxicity could theoretically be achieved by integrating the transgene into a specifically targeted and safe genomic locus. The inability of current gene therapy vectors to direct transgene integrations precisely into safe genomic sites has been associated with oncogenic changes and fatal leukemias in recent clinical trials.

We hypothesized that safe and targeted gene integration could be achieved with non-viral, site-directed genome-modifying agents such as phiC31 phage integrase and zinc-finger nucleases. If genome-modified cells are shown to be safe and capable of durably secreting transgenic FVIII at therapeutically meaningful levels, they could potentially be developed as autologous bioimplants for hemophilia A treatment. Genome modification using phiC31 integrase or ZFNs could be instrumental in developing approaches that could have wider application to other monogenic and metabolic disorders.

Cord-lining epithelial cells possess several characteristics that make them ideal as cellular vehicles for transgenic protein expression. Transgene-modified CLECs could be particularly helpful as autologous implants for treating pediatric patients.

The primary aims of my project were to evaluate (**a**) the biosafety of modifying cord-lining epithelial cells with phiC31 integrase and AAVS1-ZFNs; and (**b**) the feasibility of developing genome-modified cells as bioimplants for hemophilia A treatment.

We first proposed to establish and evaluate the efficacy, accuracy and biosafety of transgene integration in primary human cells mediated by phiC31 integrase and ZFNs. The project progressed from a comprehensive *in vitro* assessment of the genomic accuracy and genome-wide effects of transgene integration to *in vivo* transplantation of FVIII transgene-expressing cells to evaluate the efficacy, durability and safety of cellular therapy. This project was executed in the following phases: (**1**) investigation of phiC31 integrase-modified CLECs; and (**2**) investigation of ZFN-modified CLECs.

The specific aims of the project were:

1. Investigation of phiC31 integrase-modified CLECs

1.1 To clone and test bioengineered human FVIII cDNA expression from different promoters

1.2 To integrate human FVIII cDNA into genome of CLECs using phiC31 integrase and donor plasmid

1.2 To document durability of FVIII secretion from genome-modified cells

1.3 To implant transgenic FVIII-secreting cells into hemophilic mice and demonstrate phenotypic correction

1.4 To identify transgene integration sites in modified CLECs

1.5 To perform a comprehensive genotoxicity study to evaluate effects of transgene integration on transgene-modified cells. More specifically, we investigated the effects on: (**a**) the transcriptome; (**b**) whole genome copy number profile; (**c**) chromosome structures by spectral karyotyping; (**d**) *in vitro* proliferation; and (**e**) *in vivo* tumorigenicity of transgene-modified cells.

1.6 To clone and evaluate hybrid FVIII cDNA expression

1.7 To investigate the biosafety profile of selected oligoclonal cells integrated with hybrid FVIII cDNA at chromosome 8p22, specifically to document: (**a**) transgene integration into chromosome 8p22 by integration junction PCR and fluorescence *in situ* hybridization (FISH); (**b**) effects on *DLC1* expression at 8p22 by RT-PCR; (**c**) effects on the transcriptome; (**d**) effects on whole genome copy number profile; and (**e**) durability of FVIII expression and *in vivo* tumorigenicity of transgene-modified cells by implanting into immunocompromised mice.

1.8 To test feasibility of phiC31 integrase-mediated integration of hybrid FVIII cDNA in other primary human cell types and to monitor durability of FVIII expression.

2. Investigation of ZFN modified CLECs

2.1 To clone and express AAVS1 ZFNs in primary human cells

2.2 To demonstrate ZFN-mediated site-specific genomic cleavage

2.3 To optimize conditions for expression of AAVS1 ZFNs

2.4 To determine size limitations of donor DNA integration into the AAVS1 site by AAVS1 ZFNs

2.5 To evaluate the accuracy of ZFN-mediated transgene integration in CLECs

2.6 To investigate potential off-target genome modifications mediated by AAVS1 ZFNs

2.7   To integrate hybrid FVIII cDNA into the AAVS1 site in CLECs and to evaluate accuracy and efficiency of site-specific integration

2.8   To evaluate effects of transgene integration on (**a**) endogenous *PPP1R12C* gene expression; (**b**) whole transcriptome; (**c**) chromosomal karyotype

2.9   To test durability of FVIII secretion *in vitro* from modified cells with AAVS1 site-specific transgene integrations

2.10   To test and compare ZFN-mediated genome modification in other primary human cell types

Chapter 2
Results and Discussion

2.1. PhiC31 bacteriophage integrase modification of cells

Progress towards developing cell-based therapy for hemophilia requires stringent evidence of biosafety, such as lack of tumorigenicity and genotoxicity of genome-modified cells and efficacy i.e. durable FVIII expression *in vivo*, at clinically meaningful levels.

We utilized phiC31 bacteriophage integrase[1] initially to mediate stable integration of *attB*-bearing plasmid DNA constructs into the genome of CLECs. We attempted to retrieve and identify all genomic integration events to evaluate if the integrase system supports sequence-specific integration of donor DNA into a limited number of genomic sites as reported[2] and more importantly to assess potential genotoxicity of such genomic integration events. Experiments were performed to determine if genome modifications and long term *in vitro* culture adversely affected the biosafety profiles of these cells. Concurrently, we evaluated phiC31 integrase-modified CLECs for durable *in vitro* expression of human FVIII and tested them as bioimplants to correct FVIII deficiency in a murine model of hemophilia A.

2.1.1. Characterization of CLECs

CLECs are potential autologous cells for treating patients whose cords were used to derive these cells. CLECs show some stem cell characteristics such as extended proliferation, capacity to self-renew and for multilineage differentiation[3-5]. Others and we have previously been shown that CLECs express pluripotency markers such as *Oct-4*, *Nanog* and *Sox2*[3,5]. We therefore characterized CLECs derived from 5 different donors for expression of *Oct-4* and *Nanog*, two key markers of pluripotency.

Total RNA isolated from cell lysates were analyzed by reverse-transcription (RT) -PCR using gene-specific primers (**Appendix 1**) for *Oct-4* (exon-4; 159 bp) and *Nanog* (exon-5; 212 bp) mRNA. Total RNA from a human embryonic stem cell line, HUES, served as positive expression controls for both transcripts. Gel electrophoresis of RT-PCR products showed *Nanog* and *Oct-4* expression in CLECs from all 6 donor samples (**Figure 1**). Densitometric measurements of RT-PCR products were normalized to transcript levels of α-actin and compared to transcript levels in HUES. *Nanog* transcript levels were 50.4% -58.3% relative to HUES, while *Oct-4* transcript levels ranged from 46.2% - 56.4%.

Cell lysates from CLECs, HUES (positive control) and primary human fibroblasts (negative control) were immunoblotted with specific antibodies to confirm the expression of Nanog (35 kDa) and Oct-4 (45 kDa) proteins. α-Actin served as protein loading control. Nanog and Oct-4 proteins were detected in all 6 CLEC samples, albeit at lower and variable levels compared to HUES (**Figure 1**). The weaker signal for α-actin in the HUES cell lysate indicated lower total protein loading for this sample, and suggested an overall higher level of Oct-4 and Nanog expression in HUES which is known to be pluripotent, relative to CLECs[6]. By way of contrast, no pluripotency markers were detected in a differentiated cell type, primary human fibroblasts.

The lower but significant levels of Nanog and Oct-4 proteins corroborated mRNA transcript levels detected by RT-PCR in CLEC samples. Taken together, these data showed that CLECs express at least some key pluripotency genes that might explain their capacity for proliferation and multilineage differentiation.
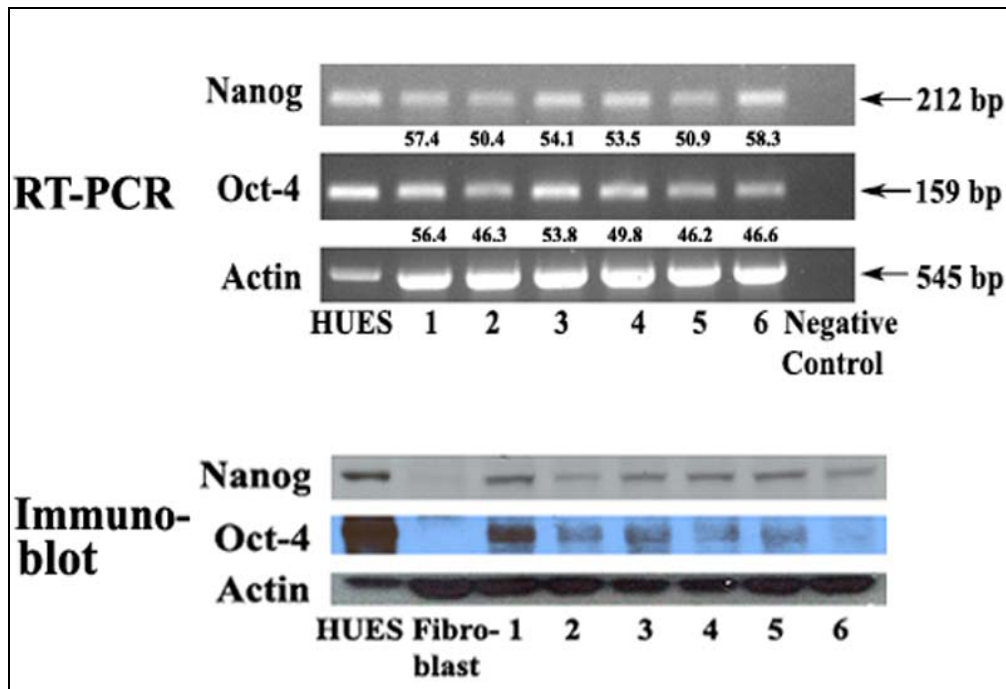
Figure 1 **Characterization of CLECs for pluripotency markers.** RT-PCR and immunoblot analysis of different CLEC samples (1-6), human embryonic stem cell line (HUES, positive control) and human primary dermal fibroblasts (negative control) for expression of the pluripotency markers, Oct-4 and Nanog and a housekeeping gene, α-actin. Negative control for RT-PCR was a minus template PCR. Shown below RT-PCR gel images are quantitative levels of Oct-4 and Nanog transcripts (normalized to actin) relative to the HUES sample.

2.1.2. Optimizing conditions for electroporating CLECs

We resorted to electroporating CLECs with plasmid DNA as lipid-based transfection reagents were inefficient (data not shown). Conditions were optimized using a BTX ECM 830 electroporator and 10 μg of pEGFP-C1 plasmid DNA. We investigated the effect of pulse voltage, pulse duration and electroporation buffer composition on transfection efficiency (percentage of GFP-positive cells) and cell viability (propidium iodide unstained cells) by flow cytometry 1 day post-electroporation.

CLECs were electroporated with pulse voltages (**Figure 2A**) from 160 V to 280 V, pulse durations (**Figure 2B**) from 10 ms to 40 ms and with different electroporation buffers (**Figure 2C**). The best setting from each experiment was that which gave the highest percentage of GFP-positive cells and least cell death. Based on these criteria, the optimal electroporation condition was determined to be a single 240 V pulse delivered for 30 ms with cells suspended in NC electroporation buffer[7]. With these settings, approximately 13% of treated cells were transfected with 89% viability. This protocol was used for all phiC31 experiments.
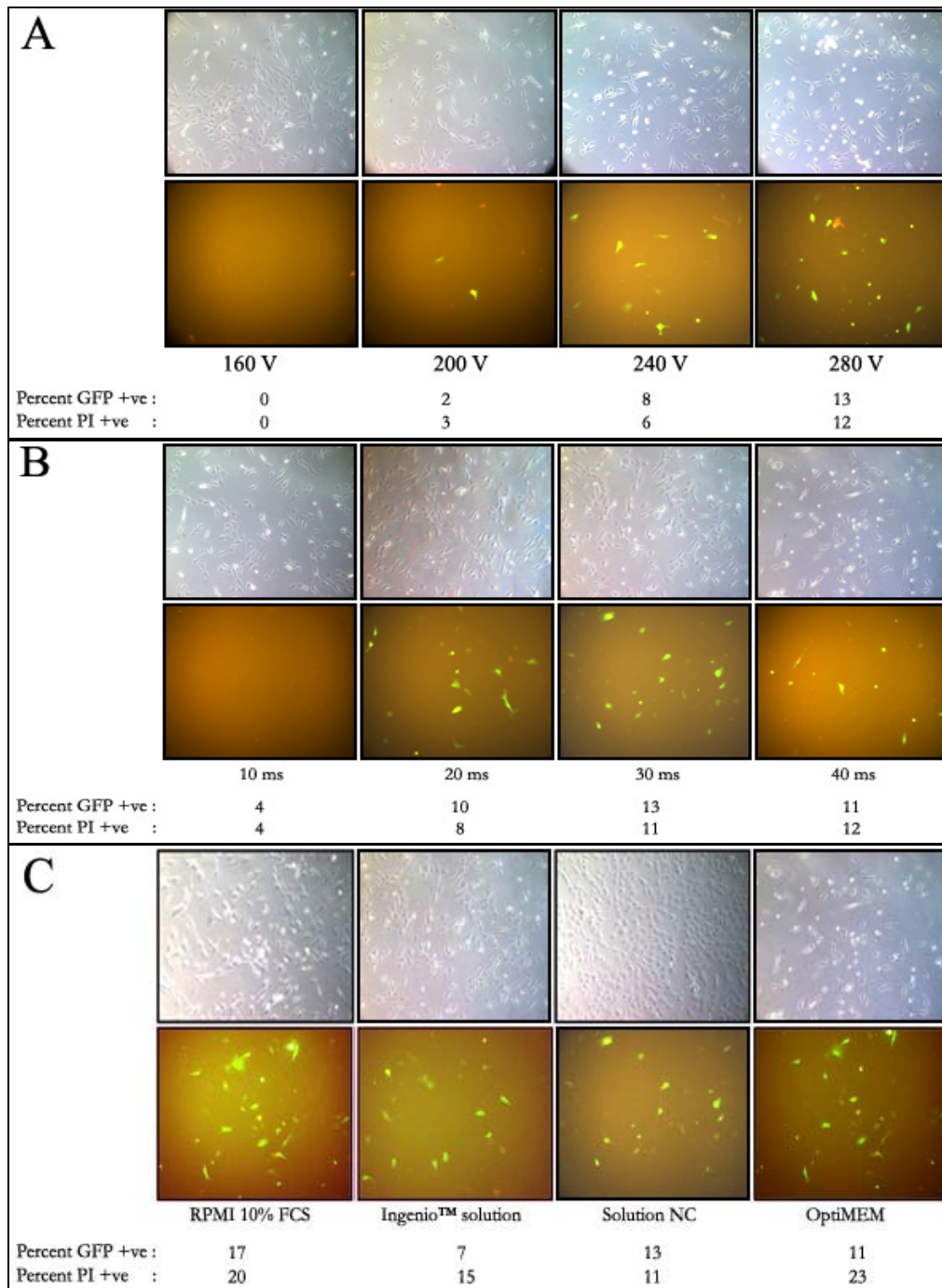
Figure 2 **Optimization of electroporation conditions.** Effect of (**A**) voltage (**B**) pulse duration and (**C**) electroporation buffer composition on gene transfer efficiency (percent GFP +ve) and cell mortality (percent propidium iodide +ve) of CLECs electroporated with 10 μg pEGFP-C1 plasmid in solution NC (unless otherwise indicated) and a single pulse delivered with a BTX electroporation device. Pulse duration was fixed at 25 ms for voltage optimization while voltage was fixed at 600 V/cm for pulse duration optimization. Electroporation buffer optimization was performed with a single 30 ms pulse at 600 V/cm. Representative brightfield and fluorescence images taken 1 day post-electroporation are shown (original magnification x100). Flow cytometric analysis for GFP expression and propidium iodide uptake was performed 1 day post-electroporation on trypsinised cells stained with 1 μg/ml propidium iodide. Data are reported as percent GFP +ve and percent PI +ve, respectively.

2.1.3. Evaluation of promoter strength on FVIII expression in CLECs

Having optimized gene transfer conditions for CLECs, we next tested their ability to secrete human FVIII *in vitro* and evaluated different promoters to drive human FVIII expression in these cells.

CLECs transfected transiently (day 3) or stably (day 15) with B-domain deleted (BDD)-human FVIII F309S plasmids expressed from simian virus 40 (SV40), elongation factor 1 alpha (EF-1α), cytomegalovirus (CMV) or human ferritin light chain (hFer) promoters were assayed for FVIII activity from overnight conditioned culture media (**Figure 3**). CLECs electroporated with pEGFP-C1 plasmid served as a negative control for FVIII expression. The percentage of GFP-positive cells by flow cytometry was $42 \pm 2.5\%$ (mean $\pm$ SEM; n=3).

For both transient (day 3) and stable (day 15) gene transfer, significantly greater levels ($P<0.05$) of FVIII were secreted from cells in which FVIII cDNA was expressed from human ferritin light chain promoter (day 3: 51.88 ± 2.38 mUnits/10^6 cells/24 hr; day 15: 350 ± 13.18 mUnits/10^6 cells/24 hr, mean $\pm$ SEM; n=3) compared to CMV (day 3: 19.29 ± 2.51 mUnits/10^6 cells/24 hr; day 15: 107.8 ± 4.92 mUnits/10^6 cells/24 hr, mean $\pm$ SEM; n=3) or EF-1α promoter (day 3: undetectable; day 15: 12.41 ± 4.12 mUnits/10^6 cells/24 hr, mean $\pm$ SEM; n=3). No FVIII activity was detected from CLECs transfected with SV40 promoter. For EF-1α, CMV and hFer promoters, higher levels of FVIII ($P<0.05$) were secreted from G418-selected CLECs stably expressing FVIII (day 15) compared to transiently transfected CLECs (day 3). This was not surprising because G418 selection enriched the culture with cells harboring the integrated FVIII transgene.

As the human ferritin light chain promoter induced the highest FVIII secretion, all subsequent experiments used the plasmid, pattB Hfer BDD-human FVIII F309S.
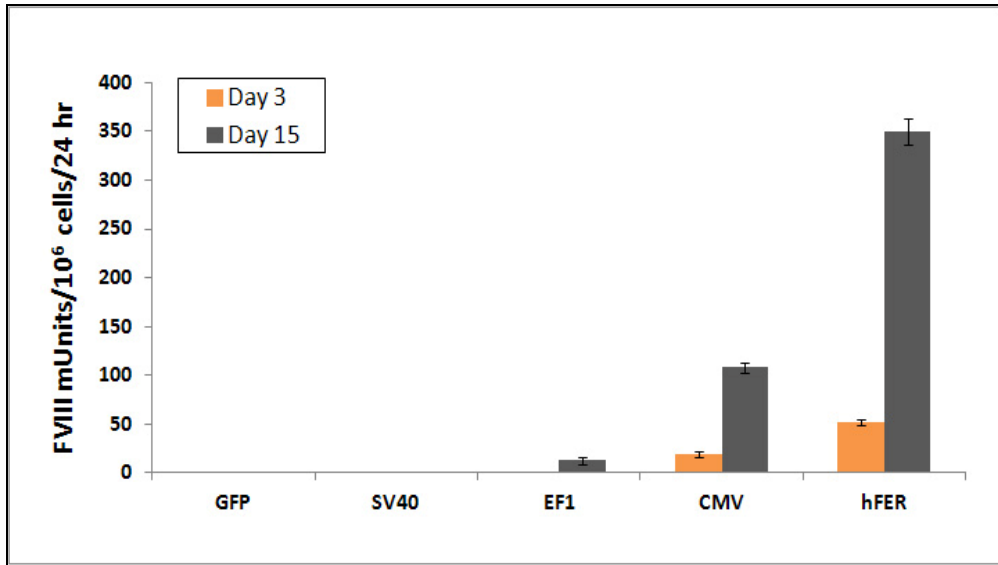
Figure 3 **Comparison of FVIII expression from different promoters in CLECs.** FVIII activity in conditioned media of unselected and G418-selected CLECs, 3 (orange bars) and 15 days (grey bars) after electroporation with a control plasmid (GFP) or BDD-human FVIII F309S cDNA expressed from SV40, EF-1α, CMV or human ferritin L promoter. Data are mean ± SEM (n=3).

2.1.4. Evaluation of integration frequency

PhiC31 integrase has been reported to mediate stable site-specific genomic integration of *attB*-bearing plasmids in a variety of cell lines and primary cells[1]. We investigated its ability to stably modify CLECs by co-electroporating an *attB*-bearing reporter plasmid, pEGFP-C1 attB, with or without a plasmid expressing phiC31 integrase. EGFP-expressing CLECs were FACS-sorted, seeded into 10 cm tissue culture dishes at initial densities of 5000 cells or 2000 cells per dish, subjected to G418 selection for 7 days after which the number of G418-resistant GFP-expressing cell colonies were enumerated using a fluorescence microscope.

From an initial seeding of 5000 cells/dish (in triplicate dishes), scoring of colonies indicative of stable genomic integration of pEGFP-C1 attB plasmid, yielded 134 ± 11 colonies (mean ± SD) when co-electroporation was performed with integrase compared to 8 ± 5 colonies when transfection was performed without integrase. Data from parallel seeding of 2000 FACS-sorted GFP-expressing CLECs (also in triplicate) showed 68 ± 12 stable colonies with integrase, and 8 ± 4 without integrase. These data showed an average integration frequency of 3.0% with integrase compared to 0.3% without integrase. **Figure 4** shows that G418 selection of integrase-modified cells resulted in a highly enriched population of GFP-expressing CLECs.
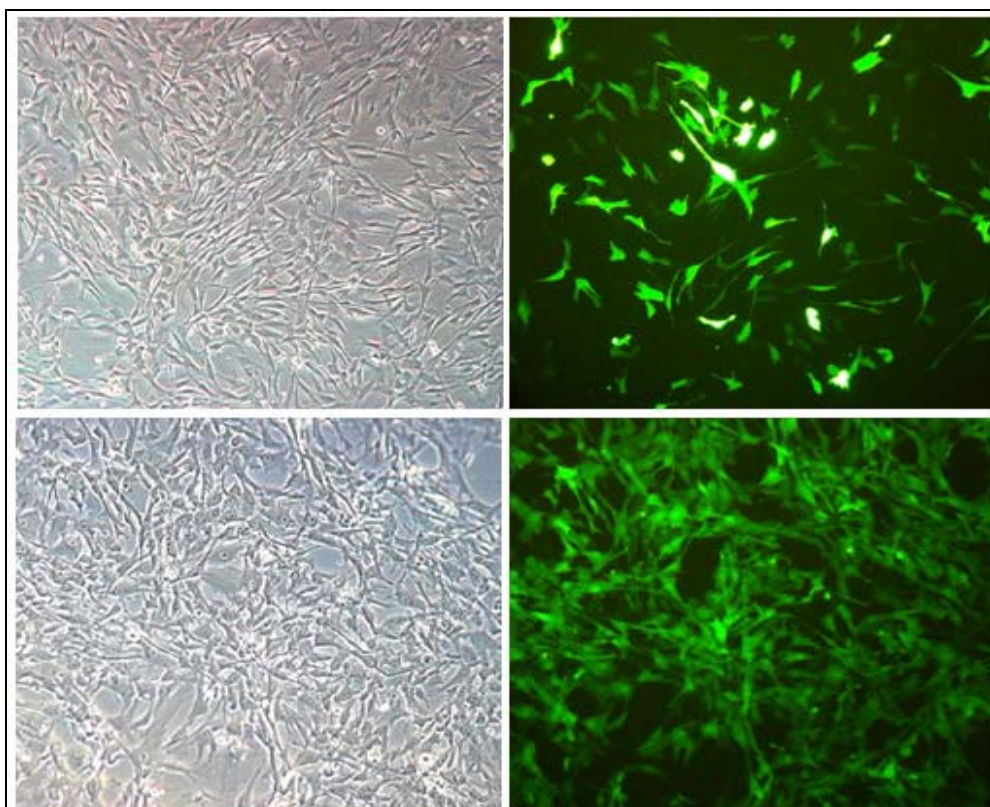
Figure 4 **PhiC31 integrase-modified CLECs stably expressing EGFP.**
(**Left**) Brightfield and (**right**) fluorescence images of CLECs co-electroporated with
10 μg of pEGFP-C1 attB and 1.5 μg of pCMV-Int phiC31 integrase plasmid DNA,
(**top**) 1 day post-electroporation and (**bottom**) 15 days after selection with G418
(original magnification x100).

2.1.5. Integration profiles of genome-modified CLECs

We next proceeded to determine the integration profiles of phiC31 integrase modified, G418-resistant, GFP-expressing CLECs stably integrated with pEGFP-C1 attB plasmid. A plasmid rescue method (detailed in section 4.3.4.1) was used to recover integration events which were then sequenced with vector specific primers to identify genomic regions at integration junctions. Workflow for identification of phiC31 integrase-mediated integrations in the genome is outlined in **Appendix 4**. Sequences and corresponding chromatograms from a single recovered integration event in chromosome 8p22 is illustrated as an example (**Appendices 4 and 5**).

We documented 44 independent integration events that mapped to 18 cytobands (**Table 1**) by sequencing 90 and 200 plasmid rescued clones from a mixed and clonal population, respectively, of genome-modified CLECs. An integration event was considered to be independent if chromosomal sequences flanking the integration site were different from all other integrations retrieved. Alignment of integration site sequences using the Multiple EM for Motif Elicitation (MEME) program revealed a shared motif among 12 cytobands (E = 5.9×10^{-10}) (**Figure 5**) that was 75% identical to a 28-base motif previously identified[2], thus reaffirming the sequence specific integrations mediated by phiC31 integrase. Integrations into the 8p22 site accounted for >40% of all integrations, confirming previous reports of frequent integrations into this locus[2, 8].
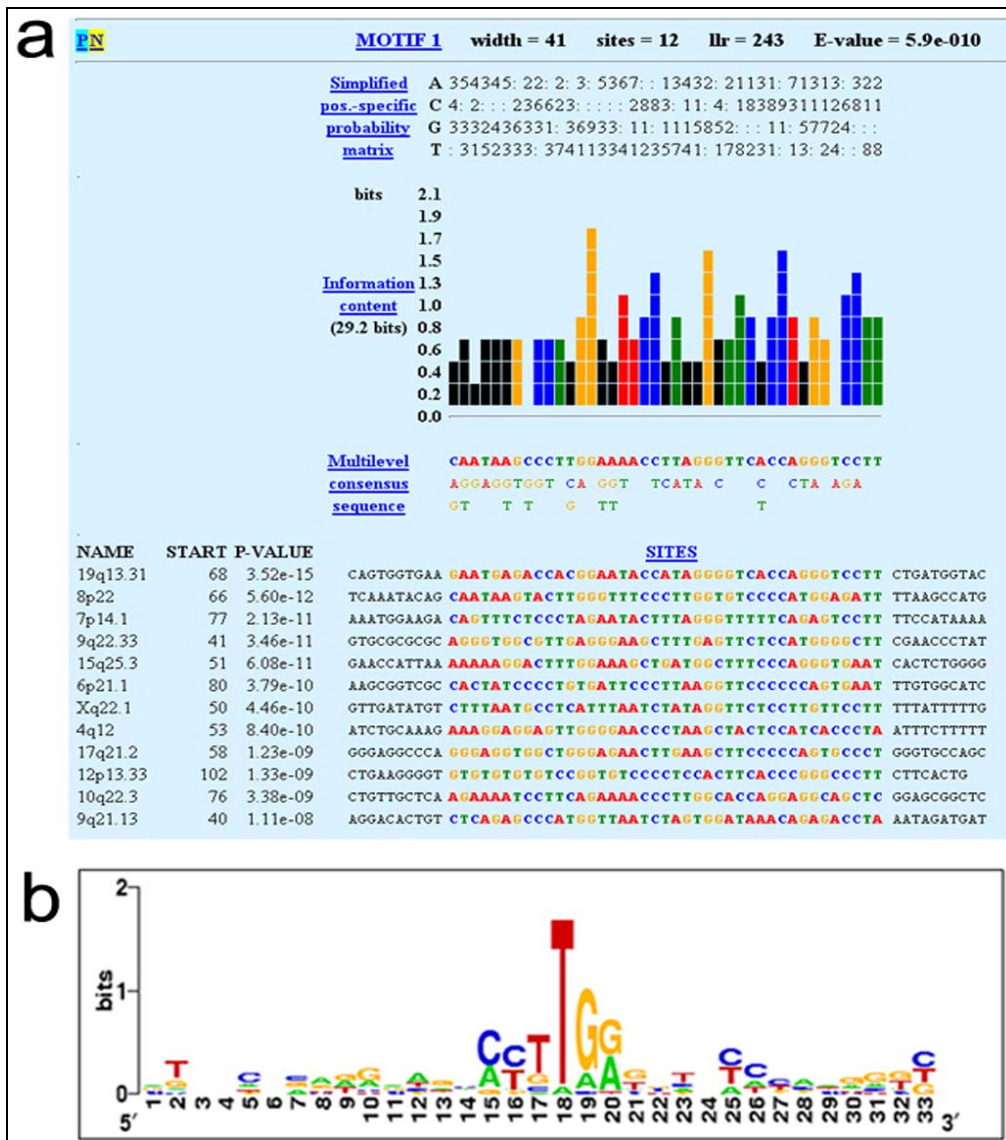
Figure 5   Site specificity of phi C31 integrase-mediated transgene integration. (a) Output of MEME motif finder from 100 bp sequences surrounding each cross-over junction identifies a common motif of endogenous *attP* sequences. A motif of 41 nucleotide bases (width) was identified among 12 aligned sequences (sites), with a log likelihood ratio (llr) of 243 and E-value of 5.9 x 10^{-10}. The simplified position-specific probability matrix scores for each nucleotide at the specified position within the motif are given (":" represents a score of 0).   The information content diagram depicts the degree of conservation of each nucleotide base at a specific motif position and reports their corresponding frequency of occurrence (measured in bits). Columns are color coded according to the majority category of nucleotides occurring at a specific location among the aligned sequences (black if no nucleotide has a frequency above 0.5). Multilevel consensus sequence shows the most conserved nucleotide at each motif position. The multilevel consensus sequence (E = 5.9 x 10^{-10}) and corresponding P-values for the 12 different integration sites are ranked in order of significance. **(b)** Sequence logo of the 33 bp sequence at cross-over junctions in CLECs (weblogo.berkeley.edu).

We next analyzed genomic site categories of phiC31 integrase-mediated integrations. Of the 44 integration events, 30 were intronic, 7 intergenic, 2 exonic and 5 within repeat elements. Most integrations were intronic (30 of 32 events) and only 2 were exonic (both were 7p14.1 integrations into exon 15 of *GLI3*). Moreover, >70% of these integrations were >50 kilobases (kb) away from transcription start sites, unlike retroviral vectors that have a predilection for integrating in close proximity to transcriptional start sites[9] where others have reported effects on altering gene expression[10]. However, our data showed that the expression of nearly all genes located within a 1 megabase (Mb) window centered on each integration site in integrase-modified CLECs was comparable to wild-type CLECs. The sole exception was a two-fold increase in *LNX1* expression, located 476 kb from the 4q12 integration site (**Table 1**). Transgene integrations close to or within known oncogenes or tumor suppressor genes raise serious concern. In our study, three independent integration events were within potential oncogenes or tumor suppressor genes (*DLC1*, *FOSL2*, and *GLI3*). Fifteen oncogenes and tumor suppressor genes were located within a 1 Mb window among 44 independent integration sites at a median distance of 224 kb (range 2–463 kb). Despite this, none of these genes showed significantly altered expression by transcriptional profiling.

Table 1   **Phi C31 integrase-mediated transgene integration sites in CLECs**

Cytoband	Number of independent integrations (clonal or mixed population)	Integration sites	Target gene (Intragenic integration)	Nearest gene (Intergenic integration)	Transcriptional effect	
					Target gene/ Nearest gene	1 Mbp Window
1p36.31	1 mixed	Intron 1	*NPH4 (1.5 kb)*		NA	No change
2p23.2	4 mixed	Intron 2	*FOSL2* (1.3 kb)		No change	No change
4q12	3 mixed	Intergenic		*CHIC2* (129 kb)	No change	*LNX1* (increased)
6p21.1	1 mixed 1 clonal	Repeat		*NCR2* (17 kb)	No change	No change
7p14.1	2 mixed	Exonic	*GLI3* (274 kb)		No change	No change
8p22	16 mixed 2 clonal	Intron 7	*DLC1* (384 kb)		No change	No change
8q24.22	1 mixed	Repeat		*NDRG1* (56 kb)	No change	No change
9q21.13	1 mixed	Intron 4	*THEM2 (28 kb)*		No change	No change
9q22.33	1 mixed	Intergenic		*C9Orf156* (0.07kb)	No change	No change
10p12.31	2 mixed	Intron 12	*DNAJC1* (247 kb)		No change	No change
10q22.3	1 clonal	Intron 1	*ZMIZ1* (2 kb)		NA	No change
10q26.11	1 mixed	Repeat		*PDZD8* (17 kb)	NE	No change
12p13.33	2 mixed	Intergenic		*FBXL14* (62 kb)	No change	No change
15q25.3	1 mixed	Intergenic		*AGBL1* (185 kb)	No change	No change

17q21.2	1 mixed	Intron 19	*JUP* (149 kb)		NE	No change
19q13.31	1 mixed	Intron 2		*ZNF 223* (7.8 kb)	No change	No change
20q11.23	1 mixed	Intron 1		LOC 128434 (15 kb)	NA	No change
Xq22.1	1 mixed	Repeat		*DRP2* (23.7 kb)	No change	No change

Vector integration sites were retrieved by plasmid rescue from genomic DNA of mixed and clonal populations of CLECs stably integrated with pattBEGFP-C1. Integration site sequences were mapped to the reference human genome sequence and their corresponding cytobands (genome.ucsc.edu). Transcriptional effects refer to the target gene for intragenic (intronic or exonic) integrations or to the nearest gene for intergenic integrations (distance from integration site to the transcription start site is indicated within parentheses). Transcriptional effects on all genes within a 1Mb window centered on each integration site are also shown. NA denotes genes for which the Affymetrix Human Genome U133 Plus 2.0 arrays did not have probe sets. NE denotes genes that were not expressed in wild type CLECs.

2.1.6. Transcriptome analysis of genome-modified CLECs

We performed microarray expression profiling experiments to compare the transcriptomes of naïve and phiC31 integrase-modified CLECs to determine if transgene integrations altered the expression profiles. Comparison of the transcriptomes of wild-type and a mixed population of stably integrated CLECs showed no difference in the expression levels of >96.5% transcripts. Of 11,947 CLEC-expressed genes, 94 (0.8%) showed increased expression and 57 (0.5%) showed decreased expression (defined for both as more than two-fold difference) in genome-modified CLECs compared to wild-type CLECs (**Table 2**). Functional annotation using DAVID (http://david.abcc.ncifcrf.gov) did not reveal significant association of these 151 dysregulated genes to specific pathways except for activation of p53 signaling (*CDK2*, *CCNB1*, and *IGFBP3*) and cell cycle (*CCNA2, CCNB1, Cdc20, TTK*) (Fisher's exact *P*-value modified for gene enrichment = 0.02) (**Figure 6**). Given that DAVID utilizes a default significant *P*-value of <0.01, a *P*-value of 0.02 for our analysis suggests that the identified KEGG pathways are not significant.

A major concern of integrating vector DNA into the genome is the potential for oncogene activation or inactivation of tumor-suppressor genes. We therefore scrutinized our list of altered genes to determine if they belonged to the categories of oncogenes or tumor suppressors. Cross-referencing the 151 transcriptionally altered genes in genome-modified CLECs identified 15 in a database of 1650 possible oncogenes and tumor suppressor genes[11] (**Table 2;** Gene symbols underlined and in italics). None of the identified 15 genes clustered to any particular cellular pathways by DAVID functional annotation analysis. Furthermore, three were tumor suppressor genes (*BRCA2*, *RAP1A*, and *TOP2A*) whose increased expression would be expected to promote cell death rather than proliferation. The remaining 12 genes were mainly involved in cell cycle regulation or cell adhesion.

Table 2   **Transcriptionally altered genes in stably integrated CLECs**

GENE ONTOLOGY CLASSIFICATIONS	OVEREXPRESSED GENES
Apoptosis	BIRC5
Cell cycle	ANLN, ASPM, *CCNA2*, *CCNB1*, CCNB2, CDCA1, *CDKN3*, CENPF, CEP55, DLG7, HCAP-G, NUSAP1, *RAP1A*, SOCS3, TCBA1, CDC20
Development	COL13A1, DDEFL1, S100A4
DNA repair	APOBEC3B, *BRCA2*, EIF4A1, NUDT1, RRM2, *TK1*, *TOP2A*
Immune response	C9orf26, PF4V1
Metabolism	ACAT2, CLN6, ENTPD4, GYS1, KYNU, PTGS1, SCD
Signal transduction	ABCB6, ADAMTS6, CNTNAP3, DEPDC1, HSPB6, HTRA4, IL11, JAK3, KIF14, KIF18A, KIF20A, LRP8, MMP19, MPP1, PACS1, PBK, PLK4, PLXNA2, PPM1K, PSD4, PSRC1, RAB7B, *RALGPS2*, TMUB1, TTK, TUBB3, ZNF236
Transcription	AARSL, ARHGAP22, FOXM1, GATA4, LASS6, LNX1, MAEL, MED12, MYOCD, PHF19
Others	*CASC5*, CLDN11, EMILIN1, FRMD4A, GLIPR1, HMMR, METT10D, NSUN4, SHCBP1, *TM4SF1*
Unknown	C21orf34, FAM83D, HEATR3, HSPCO49, KIAA1524, KIAAO101, LOC441061, LRRC17, LRRTM4, SAMD9, WDR69
GENE ONTOLOGY CLASSIFICATIONS	UNDEREXPRESSED GENES
Cell cycle	CCPG1, EPDR1, *GPNMB*, LAMP3
Development	DLK1, GPC4, GPM6B, SGCG, TSGA10

GENE ONTOLOGY CLASSIFICATIONS	UNDEREXPRESSED GENES
DNA repair	UBE2B
Immune response	ELMOD1, IL1RL1
Metabolism	CHI3L1, PTGDS
Signal transduction	ABCF2, ABCG1, ADAMTS5, B3GALT2, C20orf23, *CENTD1*, CLGN, CYP4V2, GNG7, *IGFBP3*, ITGB8, MCTP1, MFAP3L, MMP1, MMP3, MS4A6A, SLC36A1, SLC40A1, SLC03A1, SYPL2, TMEM118, TMEM148, TMTC2
Transcription	EGR2, MKX, ZFYVE21, ZNF441
Others	COL14A1, EPB41, HNT, *ICAM1*, LAMA1, SPON1
Unknown	BEX2, C10orf58, ENY2, HRASLS, KIAA1211, KIAA1450, LOC221091, RP5-875H10.1

Gene Ontology classification of genes whose expression was ≥ 2-fold different compared to wild type CLECs. Transcriptome profiling (Affymetrix Human Genome U133 Plus 2.0 array) was performed on a mixed population of CLECs one month after phiC31 integrase-mediated stable transgene integration. Potential oncogenes and tumor suppressor genes are underlined and italicized.
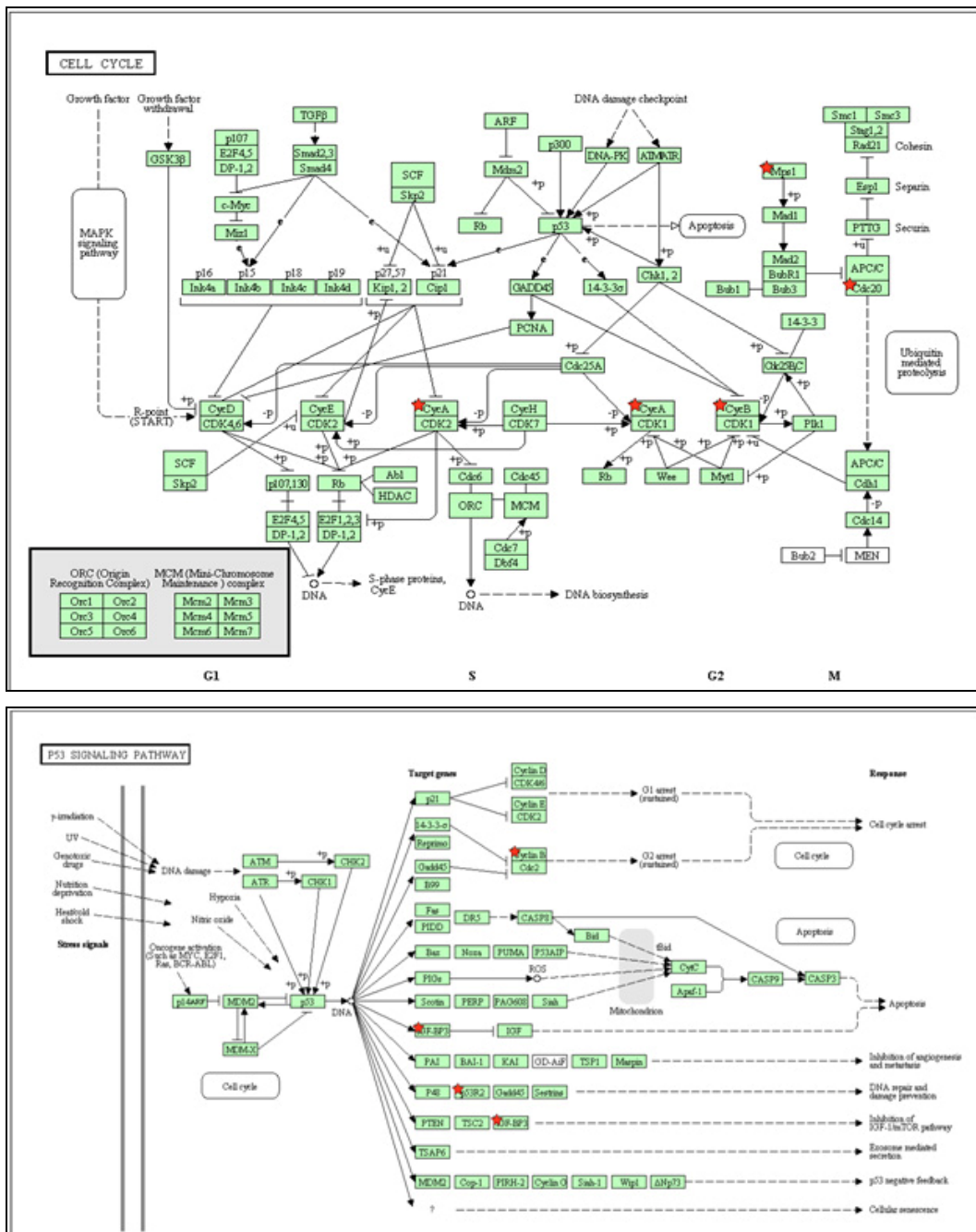
Figure 6   Functional annotation charts (KEGG pathways) of dysregulated genes identified by DAVID analysis. Dysregulated genes analyzed by DAVID functional annotation tools were identified (marked by red stars) to cluster to two main KEGG pathways. (**Top**) Cell cycle and (**bottom**) p53 pathways are depicted in these functional annotation charts. Fisher's exact *P*-value modified for gene enrichment analysis = 0.02 for both functional annotation charts suggests that the identified genes may not be significantly associated to the pathways given that the default DAVID *P*<=0.01 is considered significant.

2.1.7. Copy number change analysis of genome-modified CLECs

In order to ascertain if transgene integrations caused major amplification or deletions in the genome, we performed genome-wide copy number change analyses. High-resolution analyses of genomic DNA of naive and a mixed population of stably integrated cells showed that transgene integration into CLECs had quite minimal effects on genome copy number. Modest copy number gain was identified in two loci and copy number loss in one locus - none of which was a transgene integration site (**Figure 7A, B**). Moreover, genes residing within 1 Mb intervals centered on each copy number change region were unaltered in their expression. Further afield, we noted from our transcriptome data a 2.1- and 2.5-fold increase in the expression of *PSRC1* and *SYPL2*, respectively (**Table 2**). However, given the distance of these genes from the regions of copy number gain (**Figure 7B**), these transcriptional changes were probably independent of copy number gains. The data above are consistent with infrequent copy number changes in stably integrated CLECs, in contrast to multiple changes commonly observed in cancer cells.
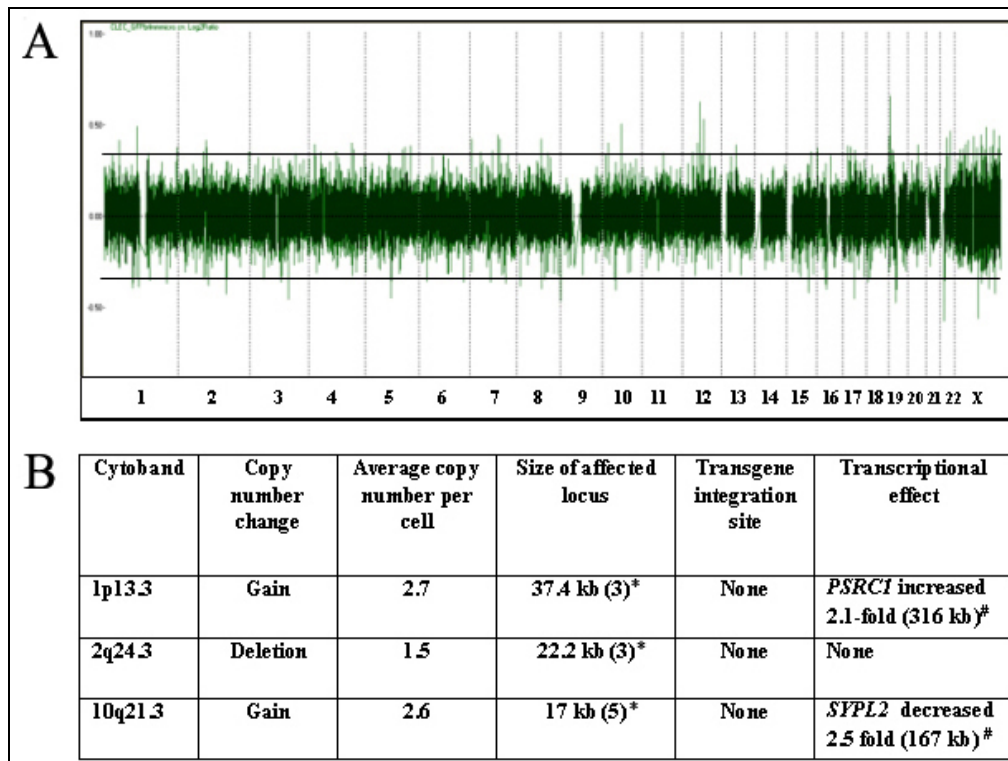
Figure 7 **High resolution copy number change analysis of phiC31 integrase-modified CLECs.** (**A**) Genome-wide copy number profile of a mixed population of CLECs stably integrated with pattBEGFP-C1 generated on Affymetrix Human Mapping 500K Array Set. Human chromosomes are shown on the horizontal axis. Log_2 signal intensity ratios are on the vertical axis. Horizontal lines are normal copy number boundaries. Dotted vertical lines demarcate individual chromosomes. (**B**) Characteristics of copy number change loci. The cytoband of genomic regions (with 3 or more consecutive probe-sets) showing significant changes in log_2 signal intensity ratios are identified, characterized as gains or deletions and indicated by the average copy number per cell. The average sizes of the affected loci (in kilobases) are listed and indicated as to whether they are transgene integration sites. Effects on transcription (2-fold or more difference compared to wild-type CLECs, as determined by Affymetrix HU133 plus expression profiling experiments) of genes in a 1 Mb window centered on each copy number change locus is shown. *The number of consecutive probe sets for each copy number change locus.. #Distance of gene from copy number change locus.

2.1.8. Fluorescence *in situ* hybridization (FISH)

Integrations of a large number of copies of the transgene could be beneficial where a high level of transgene expression is required. However, multiple integrations into the genome also increase the risk of adverse events as a result of integrations into undesirable loci (close to oncogenes, tumor suppressor genes, regulatory elements, fragile sites, etc.). We therefore performed interphase FISH using probes specific to the integrated vector to determine the number of integrations per cell. Examination of >200 stably modified CLECs revealed that >85% harbored either one or two integrations per cell (**Figure 8**). The technique employed did not differentiate monoallelic from biallelic integrations in cells that had ≥2 integrations. The low number of integrations per cell is advantageous as it reduces the likelihood of integration into high-risk sites and paves the way for selecting single cell clones that had integrations in safe genomic harbors.

No. of transgenes /cell	% Total cells	No. of cells
1	45	65
2	40	58
3	13	18
4	2	3
Total	100	144
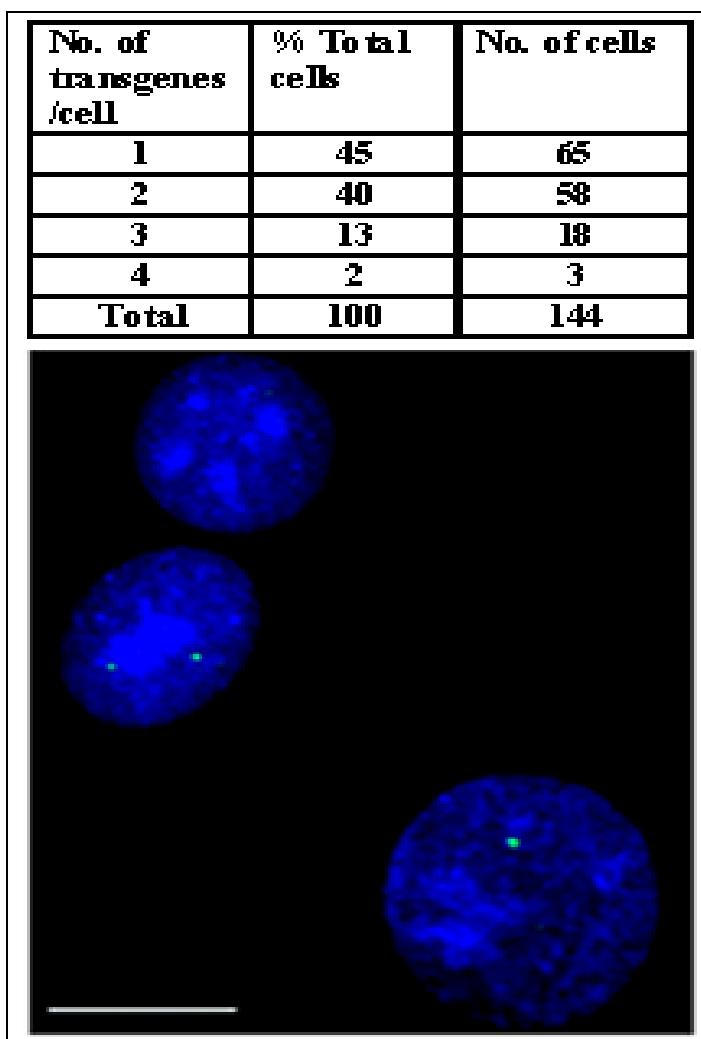
Figure 8 **Copy number of transgene integration in CLECs.** Frequency distribution of copy number of integrated transgene in CLECs determined by fluorescence *in situ* hybridization of a fluorescein-labeled vector-specific probe. Representative image of integrated transgene (green signals) in DAPI-stained interphase nuclei of CLECs. Original magnification ×1000. Bar = 5 μm.

146

2.1.9. Karyotype and spectral karyotype analyses

We spectrally karyotyped mixed populations of naive and stably integrated CLECs to determine whether phiC31 integrase induced chromosomal rearrangements. Although no chromosomal translocations or aneuploidy was detected in naive cells (40 metaphases), four of 90 metaphases from a mixed population of stably integrated cells had chromosomal translocations. Two translocations were observed only once [46XX *t*(7:13) (p21:q22); 46XX *t*(1:19) (q25, q13.3)]. A third translocation was observed twice [46XX *t*(1:18) (q25, q12)] (**Figure 9**). These led us to further analyze eight clonal populations of stably integrated cells in which we found no structural or numerical chromosomal abnormalities in >210 metaphases. The presence of nonrecurrent translocations is consistent with the known low background of chromosome aberrations in normal human somatic cells[12, 13]. That the mixed population of stably modified CLECs showed no evidence of clonal expansion of cells harboring translocations also indicated that such affected cells, when present, had no cellular growth or survival advantage.
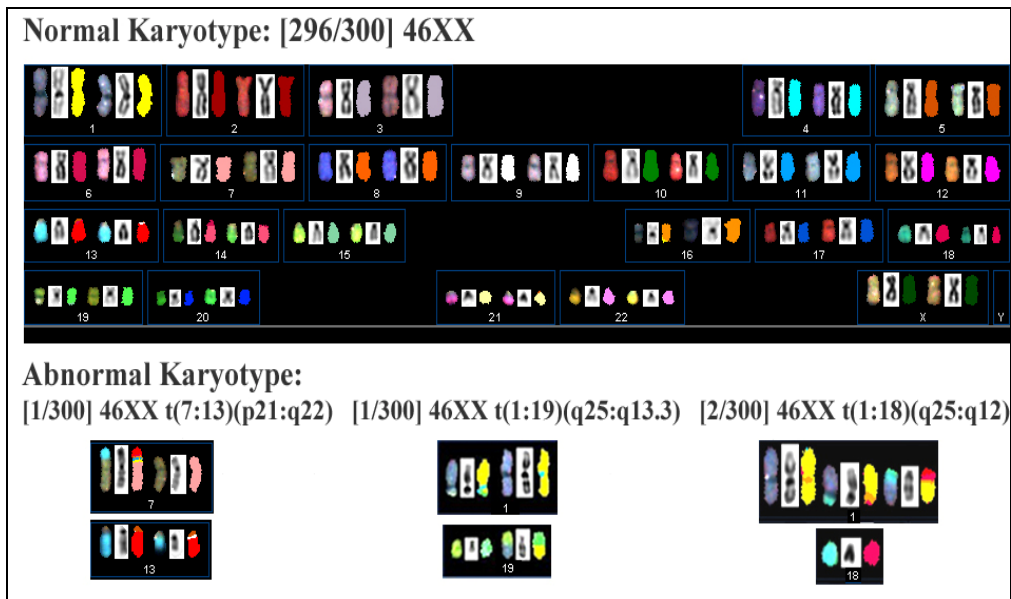
Figure 9 **Spectral karyotyping of phiC31 integrase-modified CLECs.** (**Upper panel**) Normal spectral karyotype in 296 of 300 metaphases of integrase-modified CLECs. (**Lower panel**) Rare chromosomal translocations, *t*(7:13)(p21:q22) and *t*(1:19) (q25, q13.3), were each observed in 1 of 300 metaphases. A third translocation, *t*(1:18) (q25, q12), was observed in 2 of 300 metaphases.

2.1.10. Tumorigenicity potential of genome-modified CLECs

2.1.10.1.　　　　*In vitro* colony formation

Oncogenic cells usually have increased *in vitro* proliferation capabilities and thus we investigated for this possibility by performing a colony formation assay. The number of colonies formed from an initial seeding of 100 or 200 untreated naïve CLECs and CLECs stably integrated with GFP were counted following crystal violet staining (**Figure 10**). PhiC31 integrase–mediated transgene integration did not alter the proliferative behavior of CLECs as assessed by *in vitro* colony forming assays [From initial seeding of 100 cells; 26.0 ± 0.5 (wild-type) *versus* 25.7 ± 0.3 (transgene-integrated) colonies ($P = 0.643$); From initial seeding of 200 cells; 37.0 ± 4.6 (wild-type) *versus* 38.67 ± 1.9 (transgene- integrated) colonies ($P= 0.753$), data are mean colony counts and standard error of the mean of triplicates].
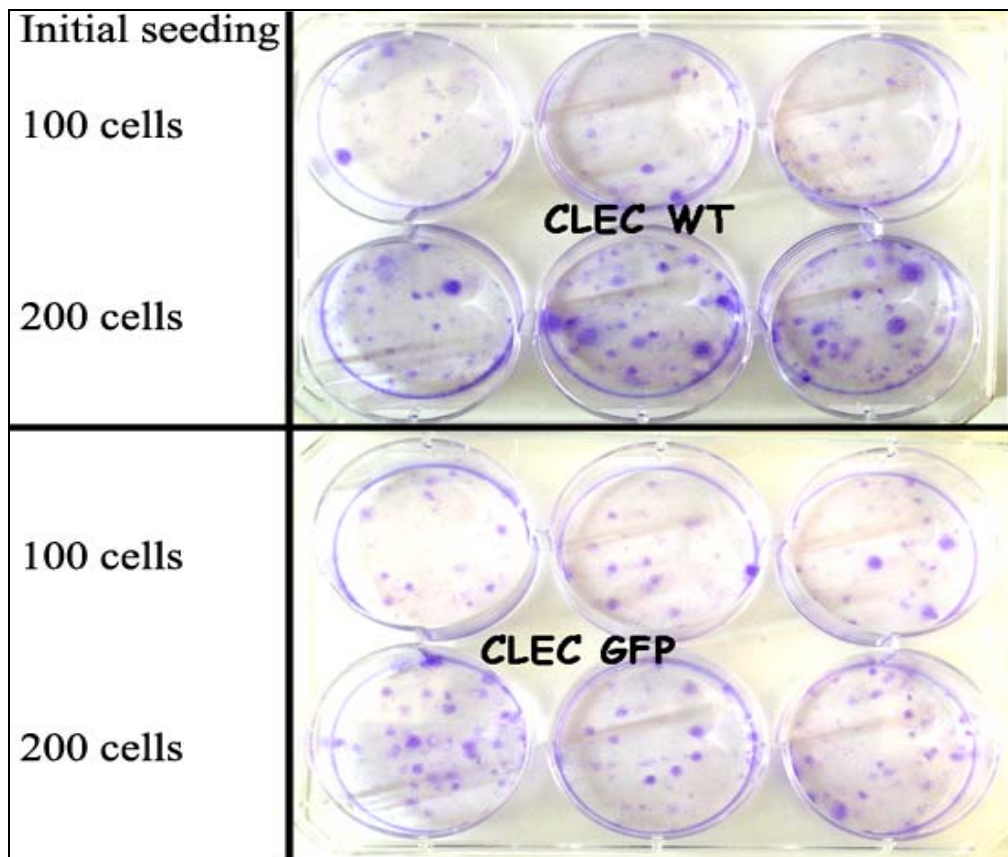
Figure 10 **Proliferation capacity by *in vitro* colony formation assay.** Crystal violet staining of cell colonies 14 days after initial seeding (in triplicate wells) with either 100 or 200 (**top; CLEC WT**) naive untreated CLECs or (**bottom; CLEC GFP**) phiC31 integrase modified CLECs stably integrated with EGFP transgene. Images of culture dishes containing crystal violet stained cells. For each seeding density, the number of colonies were enumerated and expressed as mean colony counts ± SEM.

2.1.10.2. *In vivo* implantation in immuno-compromised mice

The innate oncogenic activity of integrase-modified CLECs was evaluated by implantation into the nuchal subcutaneous region ($n = 6$) and renal subcapsular space ($n = 4$) of NOD-SCID mice. To exclude death of implanted cells, we demonstrated the survival of implanted cells by immunochistochemical staining for human cells from excised tissues at implantation sites (**Figure 11A-D**) and derived secondary cultures of the implanted cells recovered from excised implants which were subsequently identified by immunohistochemical staining (**Figure 11D**). Despite engraftment of viable CLECs at the implantation sites, no tumors developed in any of the implanted mice that were monitored for up to 4 months.
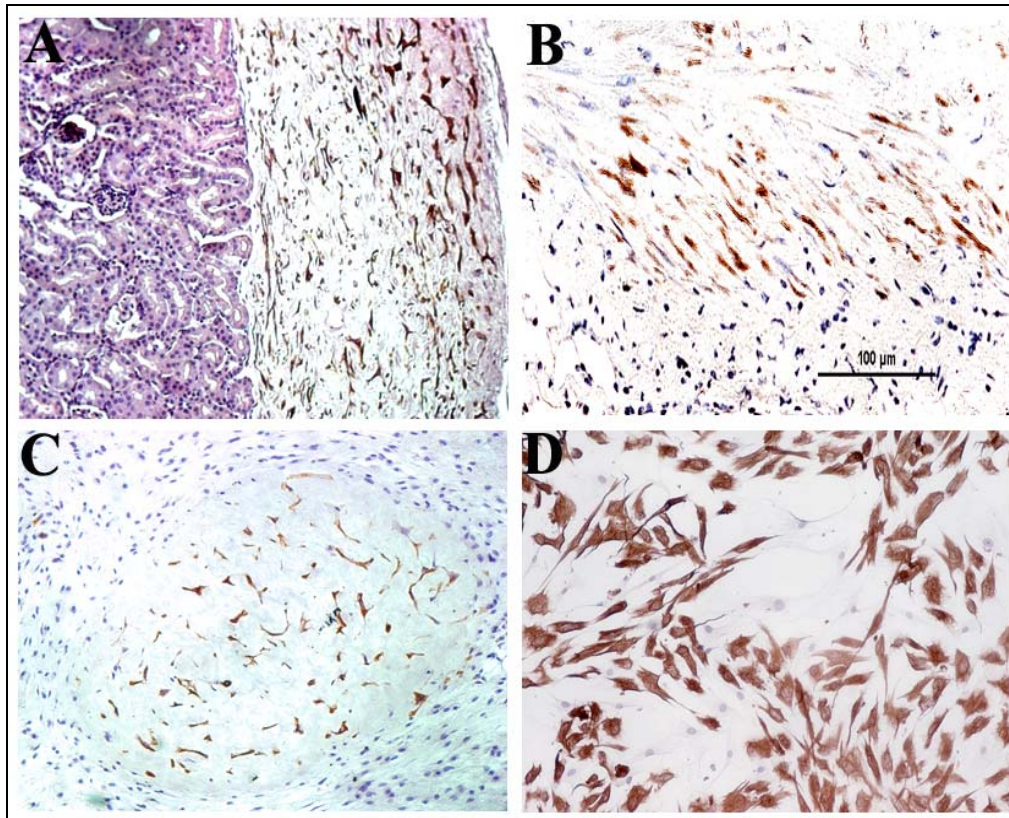
Figure 11 **Immunohistochemical staining of engrafted phiC31 integrase–modified CLECs.** Immunostaining with antihuman vimentin antibody shows renal subcapsular engraftment of CLECs (**A**) 1 month and (**B**) 3 months after implantation in NOD-SCID mice. (**C**) Engraftment of CLECs, 1 month after injection into nuchal subcutaneous region. Original magnification ×200. (**D**) Immunostaining of cultured CLECs from explants recovered from subcutaneous regions of mice 1 month after implantation. Original magnification ×100. Bar = 100 μm.

2.1.11. FVIII secretion and phenotypic correction of hemophilic mice

We co-electroporated CLECs with pattB HFer BDD-human FVIII F309S and pCMV-Int to derive FVIII-expressing cells. FVIII was readily detectable in conditioned media on day 3 (51.88 ± 2.38 mUnits per 1×10^6 cells per 24 hours) and day 15 (350 ± 13.183 mUnits per 1×10^6 cells per 24 hours; mean $\pm$ SEM; $n = 3$) after electroporation (**Figure 3**). CLECs stably integrated with pattB HFer BDD-human FVIII F309S secreted FVIII unabated for at least 5 weeks *in vitro* (**Figure 12**) whereas FVIII secretion from CLECs transfected with the same donor plasmid but without pCMV-Int never lasted >1 week (data not shown).
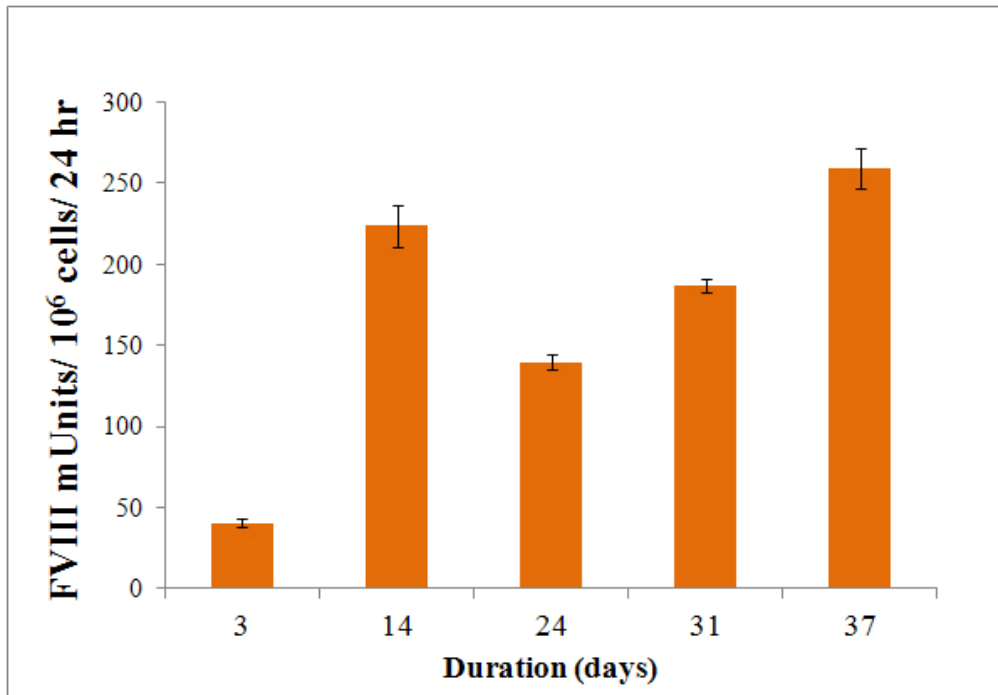
Figure 12   Durability of FVIII expression in phiC31 integrase-modified CLECs. FVIII activity (mUnits per 10^6 cells per 24 hours) in conditioned media of unselected (day 3) and G418 selected CLECs (days 14 to 37) that had been electroporated with BDD-human FVIII F309S cDNA expressed from human ferritin L promoter on day 0. Data are mean ± SEM (n=3).

In vivo secretion of transgenic FVIII was shown when subcutaneous implantation of 8×10^6 Matrigel-encapsulated, stably integrated CLECs significantly raised plasma FVIII antigen levels of hemophilic mice from 0.21 ± 0.06 % to 3.27 ± 0.56 % (mean $\pm$ SEM; $n = 5$) 3 days after implantation ($P = 0.002$ compared to control FVIII deficient mice) (**Figure 13A**). These levels significantly improved the bleeding phenotype. Control hemophilic mice implanted with naive CLECs had a mean blood loss of 797 ± 89 mg in the tail-clip assay, whereas hemophilic mice implanted with unencapsulated or encapsulated CLECs stably integrated with FVIII cDNA lost 418 ± 43 mg and 363 ± 28 mg of blood ($P = 0.03$ and 0.001, respectively as compared to control FVIII-deficient mice) (**Figure 13B**).
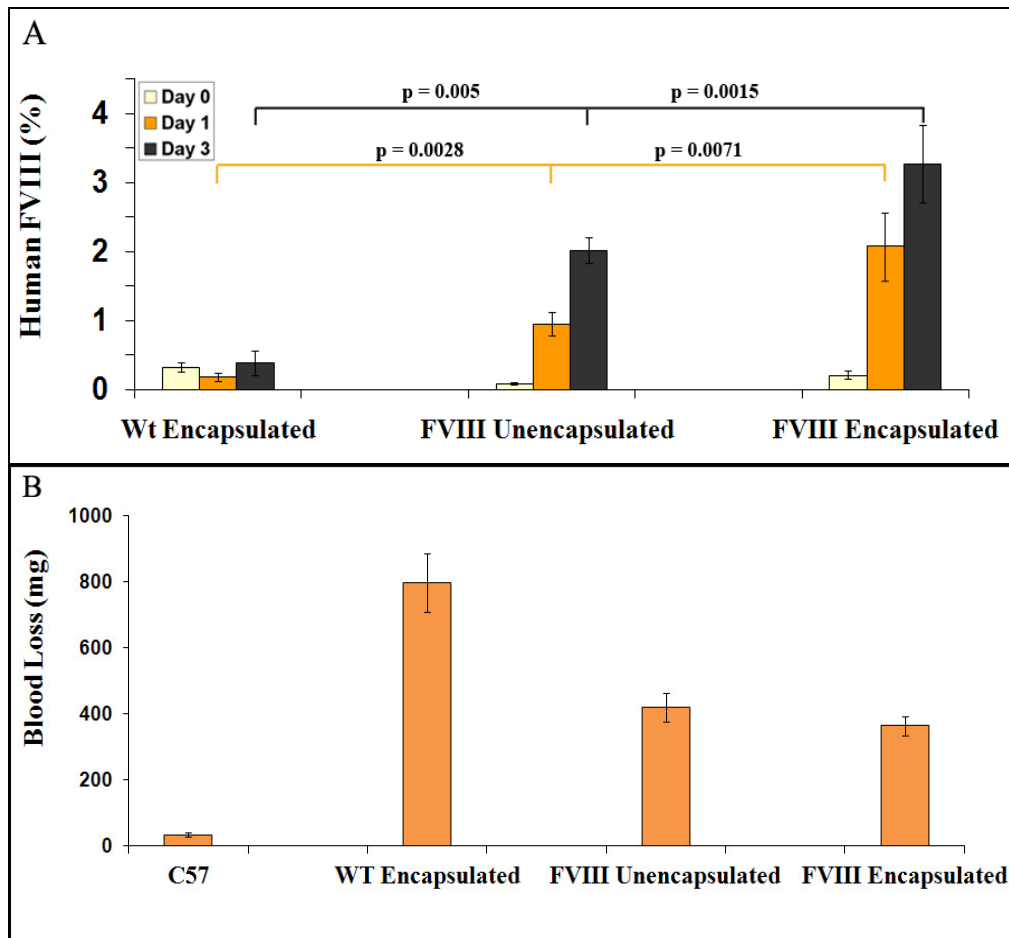
Figure 13   Detection of transgenic human FVIII in murine plasma and correction of bleeding phenotype following xenoimplantation of FVIII-secreting CLECS. (**A**) Plasma FVIII antigen levels of hemophilic mice implanted subcutaneously with 8×10^6 stably integrated FVIII-secreting CLECs that were either unencapsulated or encapsulated with Matrigel. Control animals were implanted with Matrigel-encapsulated wild-type CLECs. Plasma FVIII antigen levels were measured using an ELISA technique specific for human FVIII. FVIII levels of treated hemophilic mice were significantly higher ($P < 0.05$) on days 1 (orange bars) and 3 (black bars) compared to day 0 values (yellow bars). Data are mean ± SEM; $n = 5$ per group. (**B**) Assessment of bleeding phenotype correction**.** The amount of blood loss during a 15 minute period following a tail clip was determined for FVIII-replete C57BL/6 mice, hemophilic mice implanted subcutaneously with Matrigel-encapsulated wild-type CLECs, hemophilic mice implanted with unencapsulated or Matrigel-encapsulated FVIII-secreting CLECs. Blood loss was significantly reduced ($P<0.05$) in treated mice compared to control hemophilic animals. Data are mean ± SEM; $n = 5$ per group.

2.1.12. References

1. Calos MP. The phiC31 integrase system for gene therapy. Curr Gene Ther. (2006) 6:633-645.

2. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B*, et al*. Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol. (2006) 357:28-48.

3. Reza HM, Ng BY, Gimeno FL, Phan TT*, et al*. Umbilical cord lining stem cells as a novel and promising source for ocular surface regeneration. Stem Cell Rev. (2011) 7:935-947.

4. Zhou Y, Gan SU, Lin G, Lim YT*, et al*. Characterization of Human Umbilical Cord Lining Derived Epithelial Cells and Transplantation Potential. Cell Transplant. (2011)

5. Huang L, Wong YP, Gu H, Cai YJ*, et al*. Stem cell-like properties of human umbilical cord lining epithelial cells and the potential for epidermal reconstitution. Cytotherapy. (2011) 13:145-155.

6. Pan G, Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. Cell Res. (2007) 17:42-49.

7. Chen NKF, Wong JS, Kee IHC, Lai SH*, et al*. Nonvirally modified autologous primary hepatocytes correct diabetes and prevent target organ injury in a large preclinical model. PloS One. (2008) 3:e1734.

8. Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q*, et al*. PhiC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. Hum Gene Ther. (2003) 14:923-928.

9. Wu X, Li Y, Crise B, Burgess SM*, et al*. Transcription start regions in the human genome are favored targets for MLV integration. Science. (2003) 300:1749-1751.

10. Recchia A, Bonini C, Magnani Z, Urbinati F*, et al*. Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. Proc Natl Acad Sci U S A. (2006) 103:1457-1462.

11. Wang GP, Garrigue A, Ciuffi A, Ronen K*, et al*. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res. (2008) 36:e49.

12. Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. Am J Hum Genet. (1991) 49:995-1013.

13. Giardino D, Corti C, Ballarati L, Colombo D*, et al*. De novo balanced chromosome rearrangements in prenatal diagnosis. Prenat Diagn. (2009) 29:257-265.

2.2. Evaluation of oligoclonal CLECs with hybrid FVIII cDNA integration in chromosome 8p22

We sought to make three major improvements to refine our study of CLECs that had been genome-modified using phiC31 integrase, First, we sought to improve FVIII expression from genome-modified cells by expressing a hybrid FVIII which others have shown to be more efficiently secreted[1]. Second, we attempted to improve survival of xenogeneic human cells by implanting them into a more severely immunodeficient murine strain, the NSG mice. Third, we hypothesized that it might be possible to screen and characterize a number of genome-modified clonal CLECs that could prove to be safe for clinical cell therapy. To this end, we specifically investigated clonal CLECs with transgene integration at chromosome 8p22.

2.2.1. Hybrid human-porcine FVIII is expressed at higher levels in CLECs

We have previously shown secretion of a B-domain deleted (BDD) human FVIII variant in transfected CLECs. In an attempt to increase the level of secreted FVIII, we designed and constructed a hybrid FVIII cDNA that was similar to a construct reported to induce 10-fold higher FVIII expression[2]. We assembled a plasmid encoding a BDD-hybrid FVIII cDNA comprising A1 and A3 domains of porcine FVIII, and A2, C1 and C2 domains of human FVIII cDNA. In order to compare the efficacy of the human/porcine hybrid FVIII construct, CLECs were co-electroporated with an EGFP plasmid and a plasmid encoding either the BDD-human FVIII F309S cDNA or BDD-human/porcine hybrid FVIII cDNA, both driven from the human ferritin light chain promoter. FVIII secretion from cells electroporated with either construct was determined by chromogenic FVIII assay while transfection efficiency was determined by the percentage EGFP-positive cells measured by flow cytometry. Comparison of FVIII levels normalized to 1% of GFP-positive cells revealed an approximately 5.5-fold higher level of secreted FVIII (P=0.003) (**Figure 14**) by CLECs electroporated with the BDD-human/porcine hybrid FVIII compared with BDD-human FVIII F309S. Duplicate experiments confirmed these results although the fold-difference between hybrid FVIII and human FVIII transfected samples normalized for transfection efficiency was found to be even higher, up to 20-fold. Given the significantly superior FVIII levels achieved, all further experiments were done using the hybrid FVIII.
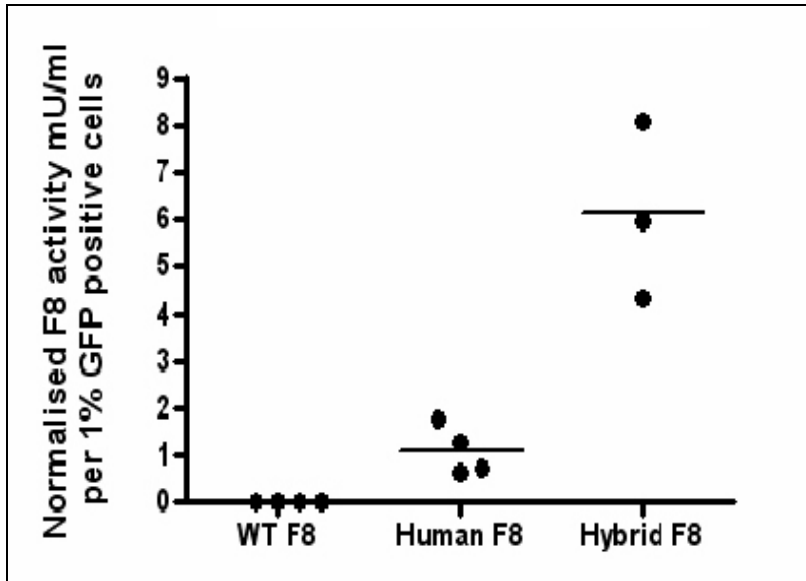
Figure 14 **Enhanced secretion of human/porcine hybrid BDD-FVIII in transfected CLECs.** FVIII activity (Coamatic® FVIII assay, Chromogenix) in conditioned media of naïve CLECs or CLECs co-electroporated with an EGFP reporter gene and either a human BDD-FVIII F309S or human/porcine hybrid BDD-FVIII. Data are expressed as mU/ml normalized to transfection efficiency. Each data point shows the mean of triplicate assays of a single experiment. An average 5.5-fold increase in FVIII levels ($P = 0.003$; Student's unpaired t-test) was detected with the hybrid FVIII cDNA compared to human FVIII cDNA.

2.2.2. Biosafety analyses of genome-modified oligoclonal CLECs

2.2.2.1. Screening for oligoclonal CLECs with 8p22 integration

Analysis of a bulk population of integrase modified CLECs suggested that there might be a small fraction of cells which acquire chromosomal aberrations as a consequence of phiC31integrase treatment. However, further analysis of 10 clonal populations of genome-modified CLECs showed that none of these analyzed clonal cells had any chromosomal abnormalities. We thus hypothesized that it might be possible to screen and characterize a number of genome-modified clonal CLECs to provide greater assurance of biosafety. These clones could then be expanded and used for *in vivo* cell therapy. Our earlier studies established that almost 45% of all integrations were single copy integrations and that close to 40% of all recovered integration events were at chromosome 8p22. Thus we sought to specifically investigate clonal CLEC populations with transgene integration at chromosome 8p22. Our main objective was to investigate whether clonal cells with FVIII transgene integration at chromosome 8p22 could prove to be safe and capable of sustaining durable FVIII expression.

A bulk population of G418-selected CLECs treated with pattB Hfer hybrid FVIII and phiC31 integrase was analyzed for durability of FVIII secretion and for evidence of transgene integration at the chromosome 8p22 locus. FVIII activity assay (**Figure 15B**) of conditioned media showed durable FVIII expression from G418-selected CLECs that were transfected with pattB Hfer hybrid FVIII and phiC31 integrase (day 6: 298.30 ± 19.01 mUnits FVIII/ 10^6 cells/ 24 hr; day 25: 377.1 ± 28.42 mUnits FVIII/ 10^6 cells/ 24 hr) but not those that were similarly transfected but not G418 selected (day 6: 298.30 ± 19.01 mUnits FVIII/ 10^6 cells/ 24 hr; day 25: 20.42 ± 4.01 mUnits FVIII/ 10^6 cells/ 24 hr), indicating that G418 selection enriched for cells with stable genomic integration of the FVIII transgene. CLECs that were transfected only with pattB Hfer hybrid FVIII and not G418 selected also showed a steep decline in FVIII levels from day 6 to day 25 (day 6: 89.02 ± 9.77 mUnits FVIII/ 10^6 cells/ 24 hr; day 25: 3.33 ± 0.71 mUnits FVIII/ 10^6 cells/ 24 hr), consistent with loss of episomal FVIII plasmid with successive cell divisions and negligible genomic integration in the absence of the integrase.

A bulk population of CLECs that was electroporated with pattB Hfer hybrid FVIII and phiC31 integrase and G418 selected was screened for transgene integration at chromosome 8p22 by junction PCR using primer pairs that recognized sequences within the FVIII vector and within a previously identified pseudo *attP* site in chromosome 8p22[3]. PCR amplification of both left and right integration junctions

provided evidence for site-specific integration of FVIII transgene cassette at the locus of interest (**Figure 15A**). Integration junction PCR products were sequenced to confirm that transgene integration had occurred at the previously identified hotspot in chromosome 8p22. Sequence analysis showed that there was a 6-bp deletion in the vector sequence and a 7-bp deletion in the genomic region at the left and right integration junctions, respectively (**Appendix 5**).
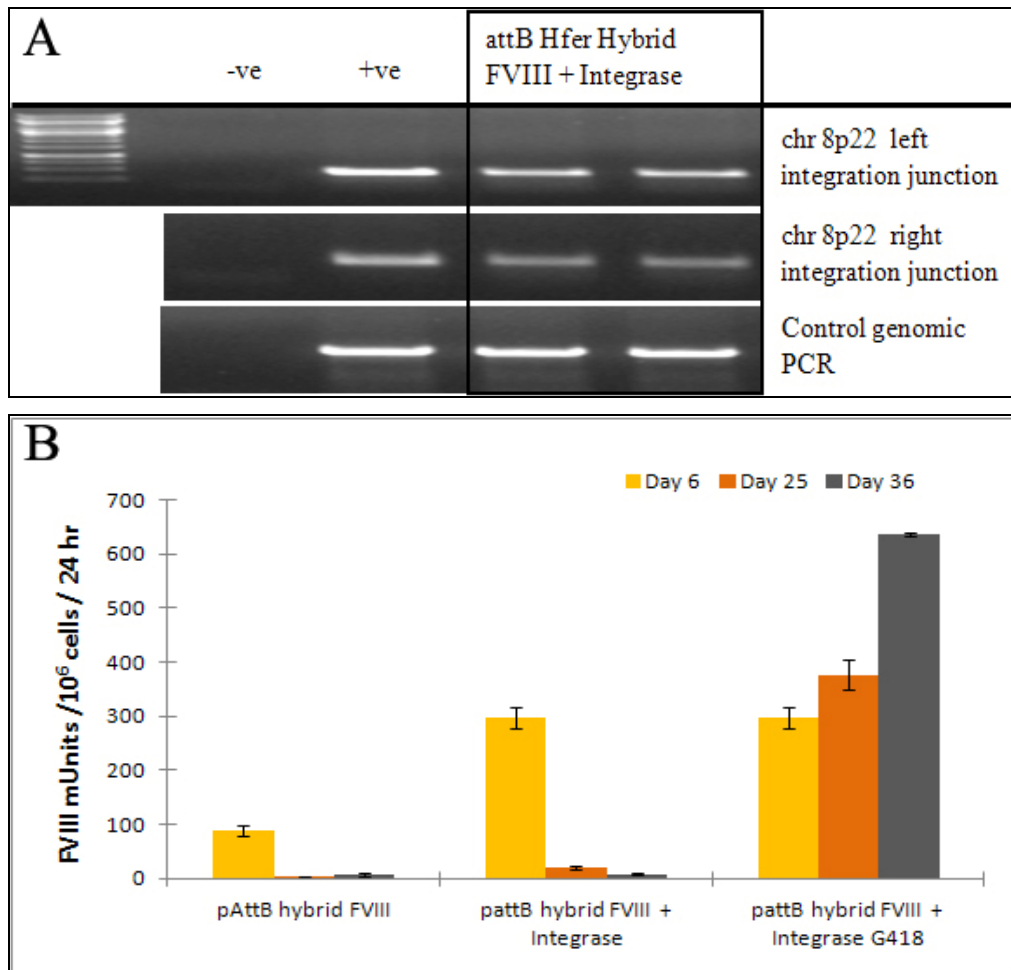
Figure 15 **Evidence of transgene integration at 8p22 locus and durable FVIII secretion from CLECs treated with pattB Hfer hybrid FVIII and phiC31 integrase.** Stable G418-resistant CLECs derived by electroporation with pattB Hfer hybrid FVIII and phiC31 integrase were screened by junction PCR for evidence of integration at chromosome 8p22 and for FVIII activity in conditioned media. (**A**) Genomic DNA (200ng) extracted from a bulk population of genome-modified CLECs was amplified with primer pairs specific for the vector and chromosome 8p22 genomic DNA to detect the presence of left and right integration junctions indicative of correct integration at the pseudo *attP* site in chromosome 8p22 (done in duplicates). "–ve" denotes minus template PCR amplification while "+ve" denotes amplification from genomic DNA isolated from a clonal line of CLEC previously identified to have integration at 8p22 locus. Control PCR amplified a 900 bp region in chromosome 19q13.42 (AAVS1 locus). Amplified products were resolved 1% agarose gel electrophoresis and imaged using BioRad® Gel Doc 2000 transilluminator and QuantityOne software. (**B**) Overnight conditioned media of genome-modified CLECs (pattB hybrid FVIII; no G418 selection, pattB hybrid FVIII + integrase; no G418 selection, pattB hybrid FVIII + integrase; G418 selection) was assayed for FVIII activity on day 6 (before G418 selection), day 25 and day 36 (after G418 selection; where indicated) post- electroporation. Data are mean ± SEM; n=3 per group.

Evidence that genome-modified cells with 8p22 integrations were present in the bulk-transfected population prompted us to derive clonal populations of genome-modified CLECs having this specific integration. We examined oligoclonal CLECs obtained by flow sorting 4 cells into each well of 96-well plates for evidence of 8p22-specific transgene integration by direct *in situ* PCR. Using primer pairs specific to the vector sequence and sequence at the 8p22 hotspot, genomic DNA from i*n situ* lysed cells were screened by PCR for presence of left and right integration junctions, indicative of complete transgene integration. A control PCR amplifying a genomic region in chromosome 19q13.42 served to verify integrity of genomic DNA and the PCR efficiency between samples. Our analysis of 72 oligoclonal populations revealed 66 that were positive for control PCR amplification indicating presence of intact genomic DNA in those samples. Of these 66 samples, 16 (24%) amplified both left and right integration junctions at the 8p22 hotspot (**Figure 16A**).

We next proceeded to test clonal CLECs with transgene integration at 8p22 for FVIII expression (**Figure 16B**). Of 13 clonal populations screened for FVIII secretion capacity, 10 secreted supraphysiological levels of FVIII (2610 – 5724 mUnits FVIII/ 10^6 cells / 24 hr), while 3 clonal cultures (#30, #18, and #47) that did not secrete FVIII were also negative for transgene integration at 8p22 (**Figure 16A**). High levels of FVIII expression detected 40 days post-electroporation support the hypothesis that transgene integration at 8p22 is compatible with durable FVIII secretion by genome-modified cells.
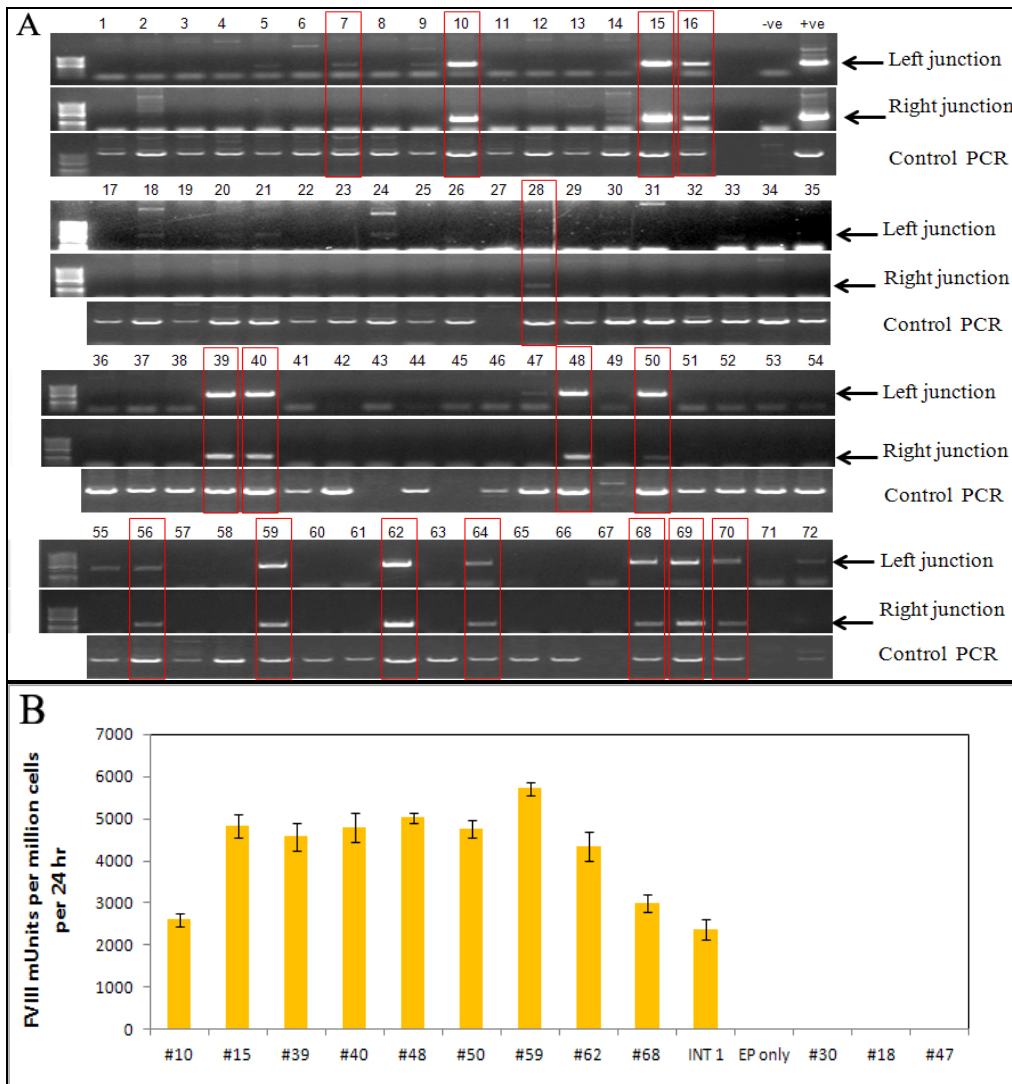
Figure 16 **Identification of oligoclonal CLECs with transgene integration at 8p22 locus and FVIII secretion.** (**A**) Oligoclonal cells (4 flow-sorted cells per well) from a bulk population of G418-selected CLECs electroporated with pattB hybrid FVIII and phiC31 integrase were investigated by direct *in situ* PCR for transgene integration at the chromosome 8p22 hotspot. *In situ* lysed cells were screened using Phusion® human specimen direct PCR kit (state manufacturer) and primers specific for vector and genomic sequences at chromosome 8p22 integration site to detect the presence of left and right integration junctions. Control genomic PCR amplified a 900 bp region in chromosome 19q13.42 (AAVS1 locus). "–ve" denotes minus template PCR amplification while "+ve" denotes amplification from genomic DNA isolated from a previously identified clonal CLEC with integration at 8p22 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Red box highlights samples which were positive for left and right integration junctions and control PCR. (**B**) CLECs that were electroporated without any plasmid DNA (EP only), CLECs from a bulk culture of phiC31 integrase-mediated integration of hybrid FVIII cDNA (INT 1) and clonal CLECs that were positive for transgene integration at 8p22 locus (# 10, 15, 16, 28, 39, 40, 50, 59, 62, 68, 69) as well as 3 clonal populations that were negative for 8p22 transgene integration (#18, 30, 47) were evaluated for FVIII secretion (day 40 post-electroporation). Overnight conditioned media were assayed for FVIII activity using a Coamatic FVIII kit. Data are mean ± SEM; n=3 per group.

We performed FISH with probes specific to chromosome 8 centromere and the integrated transgene to confirm integration at chromosome 8p22 and to determine transgene copy number in oligoclonal cells. Single copy transgene integration was noted in the vast majority of cells for each oligoclonal cell population analyzed. **Figure 17** shows the close proximity of FISH signals specific to chromosome 8 centromere and the integrated transgene could be taken as evidence of integration at chromosome 8p22, consistent with our results from junction PCR and sequencing. However, we noted that all of the oligoclonal populations screened also had cells with transgene integrations in other chromosomes. Thus, the oligoclonal populations were mixtures of cells with integrations at 8p22 and other chromosomes. We were compelled to use oligoclonal cells owing to the difficulty of deriving monoclonal cell lines by flow-sorting single cells. Given that oligoclonal cells were derived from an initial sorting of 4 cells per well and assuming all transgene integrations were single copy events, a maximum of 4 different integration sites could be expected for each oligoclonal population. Nonetheless, screening oligoclonal cells for tumorigenic potential could be considered an improvement compared to evaluating a much more heterogeneous bulk population we have shown to consist of cells with single and multiple transgene integrations and integration sites.
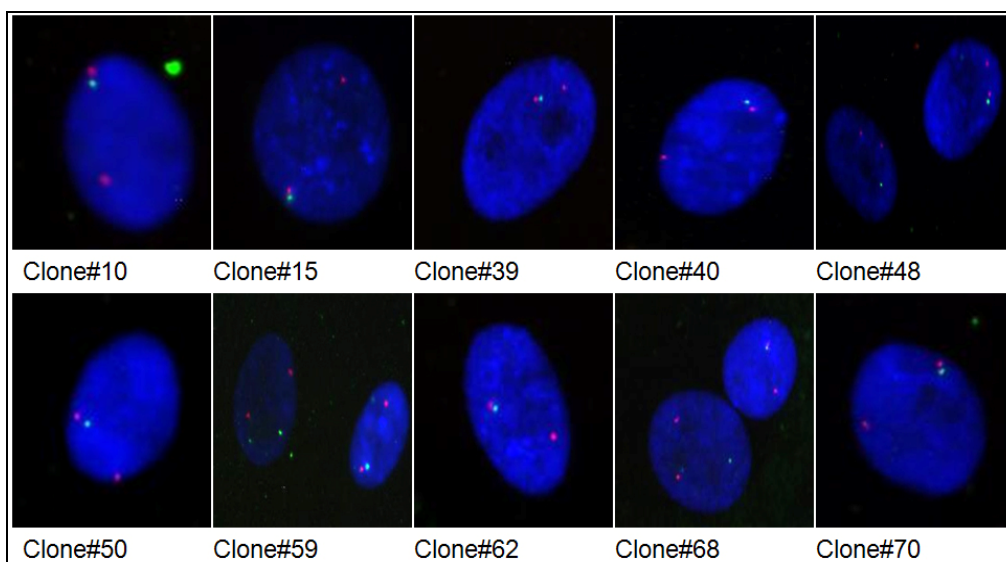
Figure 17 **FISH to verify transgene copy number and integration at chromosome 8p22 in oligoclonal cells.** Genome-modified oligoclonal CLECs identified by junction PCR to be positive for transgene integration at chromosome 8p22 were screened by FISH. Fluorescence images of DAPI-stained cells hybridized with FITC-labeled probes specific to integrated vector (green signal) and Texas Red-labeled centromeric probes specific to chromosome 8 (red signal) are shown (original magnification 600x). Integration of transgene at chromosome 8p22 is indicated by the close proximity of the vector specific green signal and chromosome 8 centromere red signal.

2.2.2.2. RT-PCR analysis of *DLC1* transcript in oligoclonal CLECs with 8p22 integration

As the transgene integration site in chromosome 8p22 is situated within an intron of a tumor suppressor gene, *DLC1*[4], it was pertinent to determine if *DLC1* expression was perturbed by the integrated transgene. We performed quantitative-RT-PCR to determine if *DLC1* transcript levels (exon 1 and exon 8) in clonal CLECs with transgene integration at 8p22 were different relative to control CLECs (electroporated without any plasmid DNA but having undergone the same number of population doublings in culture). The mean fold changes in *GAPDH*-normalized *DLC1* transcript levels in CLECs with transgene integration in 8p22 expressed relative to normalized *DLC1* transcripts levels of control CLECs are shown in **Figure 18**. Normalized *DLC1* transcript levels in clonal CLECs with 8p22 integrations ranged between 0.58 to 1.603 (exon 1) and 0.851 to 1.38 (exon 8) of normalized transcript levels in control CLECs, suggesting minimal perturbations to *DLC1* expression as a consequence of transgene integration at 8p22.
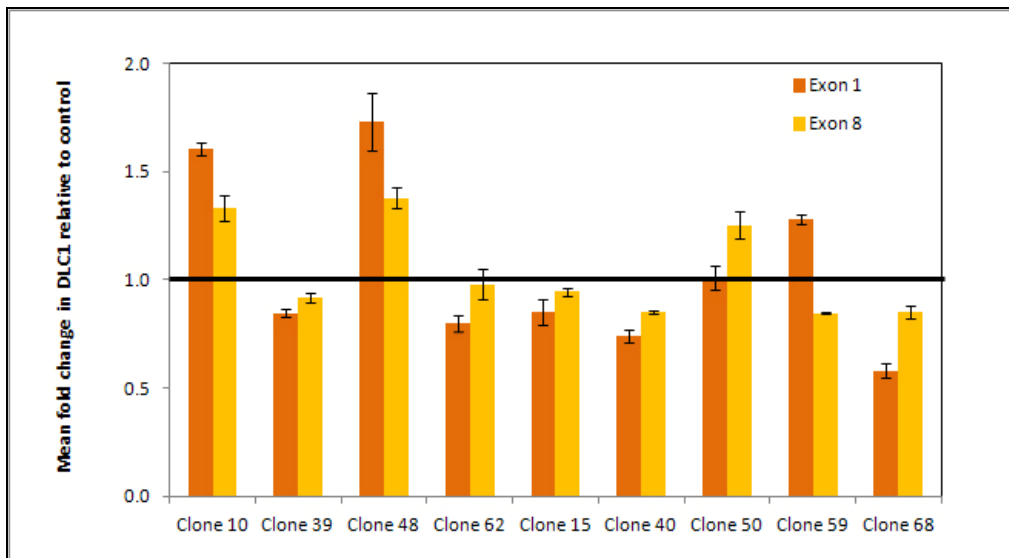
Figure 18   Quantitative RT-PCR of *DLC1* transcript levels in 8p22 oligoclonal CLECs. Quantitative RT-PCR was performed on control CLECs that had been electroporated without any DNA and oligoclonal CLECs with transgene integration at 8p22 to quantify *DLC1* transcript levels (exon 1 and exon 8) and *GAPDH* expression. The mean fold change in *GAPDH*-normalized *DLC1* expression levels in clonal samples (exon 1 and exon 8) are shown relative to that of control CLECs as determined by the $2^{-\Delta\Delta Ct}$ method[5]. Horizontal black line demarcates no change in normalized *DLC1* transcripts relative to control. A relative mean fold change in *DLC1* levels of ≥2 was considered significant. Data are mean ± SEM; n=3 experiments per group.

2.2.2.3. Transcriptome analysis of oligoclonal CLECs with 8p22 integration

The extent to which transgene integration at 8p22 altered the transcriptome of oligoclonal CLECs was investigated by transcriptome profiling on Affymetrix PrimeView expression arrays. Transcripts that differed by 2.5-fold or more compared to CLECs that were electroporated without plasmid DNA and cultured under similar conditions and length of time were identified.

Transcriptome analysis identified a total of 341 genes that were commonly altered in expression in all 8 oligoclonal CLECs with 8p22 integration compared to control CLECs (**Appendix 7**). Of these, 93 genes were up-regulated and 248 genes were down-regulated by 2.5-fold or greater. No significant changes were detected in the expression of *DLC1* gene at chromosome 8p22 or genes that mapped within a 1 Mb window (*LONRF1, KIAA456, C8orf48*) centered on 8p22 integrations for all 8 oligoclonal CLECs. Significantly altered genes were classified and clustered according to their biological functions and their proportional categorizations are depicted in **Figure 19**. Intriguingly, up to 40% of the genes with down-regulated expression were involved in cell cycle regulation, suggesting that proliferation of oligoclonal CLECs with 8p22 integration could be reduced compared to control CLECs. A potential reason for such growth retardation could be the increased metabolic burden from overexpression of a transgenic protein[6]. Given that the oligoclonal CLECs were cultured continuously *in vitro* for at least 2 months, induction of senescence associated with down-regulation of cell cycle genes could not be excluded. In either case, down-regulation of cell cycle genes and slower cell proliferation would be in direct contrast with the behavior of transformed cells.

The list of genes with altered expressed were submitted to DAVID pathway mapping analysis and compared against a list of oncogenes[7]. This clustered down-regulated genes (46 genes) mainly to cell cycle and DNA repair pathways (KEGG pathway database) (**Appendix 8**). Genes with up-regulated expression were not significantly clustered to any pathway. Comparisons of the altered genes with a list of potential oncogenes identified 9 and 31 oncogenes among the list of genes that were up-regulated and down-regulated, respectively (**Appendix 7**).

In summary, transcriptome analysis of oligoclonal CLECs with transgene integration at 8p22 locus did not bear signatures of transformed cell lines. This was evident by the fact that a large majority of cell cycle genes were down-regulated, consistent with slower growth of these genome-modified cells as opposed to accelerated growth of transformed cells.
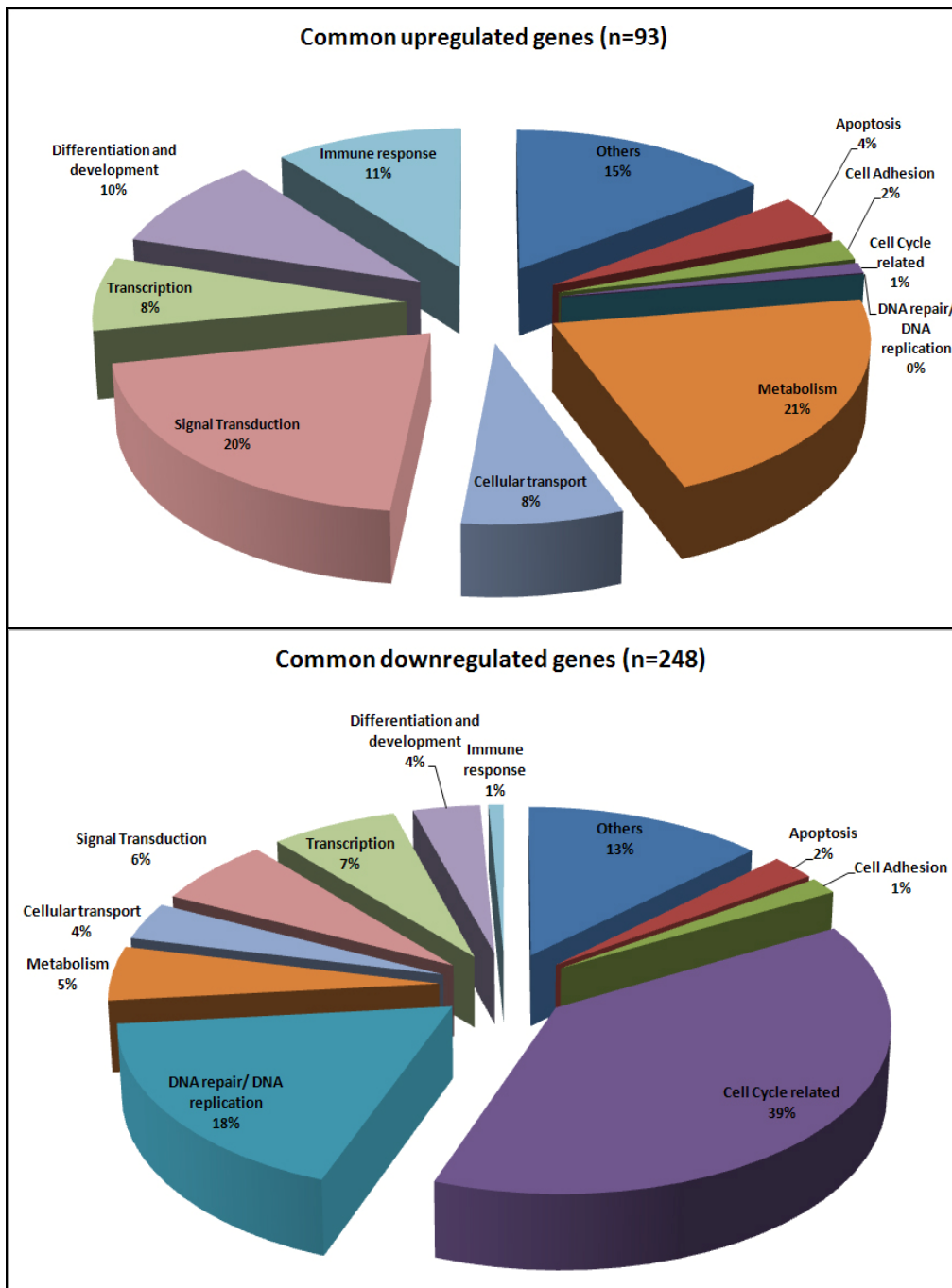
Figure 19   Classification of significantly altered genes according to biological process. Categorization of genes which differed in their expression by ≥2.5-fold in all 8 oligoclonal CLECs relative to control CLECs. The percentage of (**top**) significantly up-regulated and (**bottom**) significantly down-regulated genes categorized according to their biological functions are depicted.

2.2.2.4. Molecular cytogenetic analysis of oligoclonal CLECs with 8p22 integration

We next evaluated oligoclonal CLECs with 8p22 integrations for global copy number changes by performing a high density SNP array hybridization (Affymetrix CytoScan HD array). Our global analysis for copy number change revealed no significant change in all 8 oligoclones tested except for a single amplification event (copy number state of 3) in the peri-centromeric region of chromosome 19 (spanning 4118 kb and identified by a total of 442 probes) in clone#10 (**Figure 20**). Only a single gene, *ZNF254*, resided within this affected region but its expression was not significantly changed.
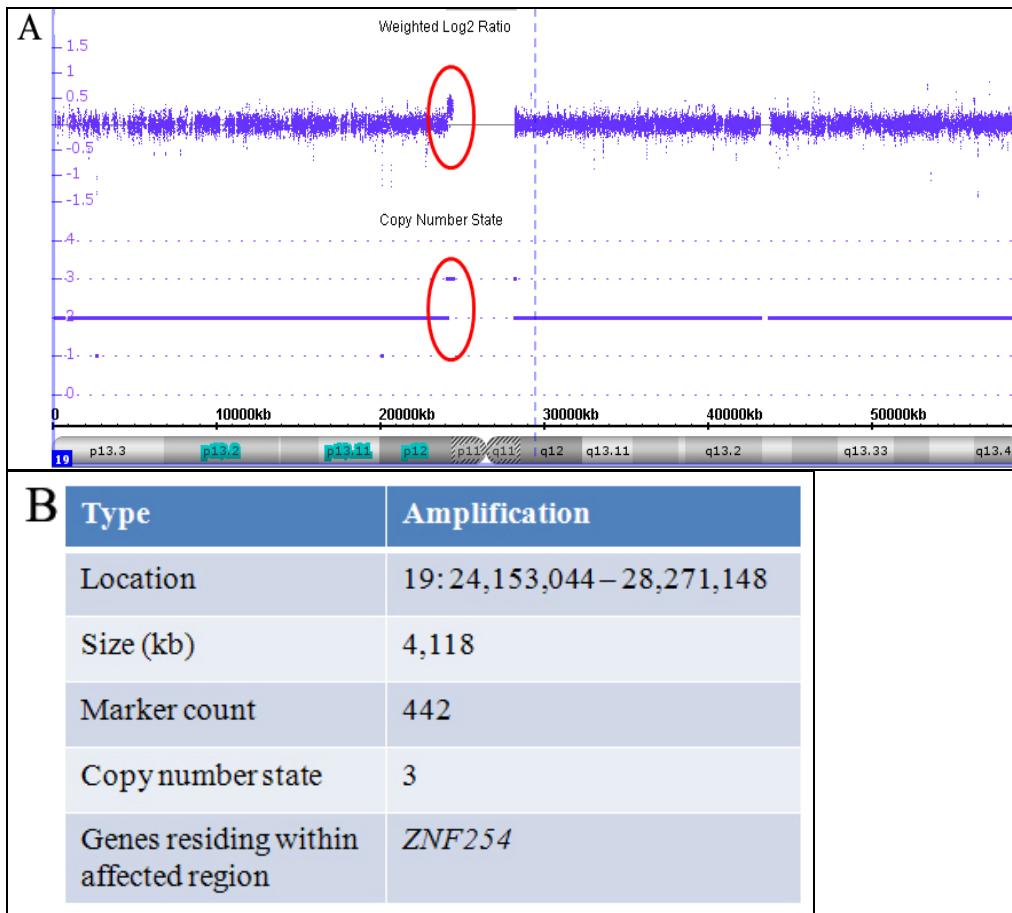
Figure 20   Copy number change observed at chromosome 19 in CLEC oligoclone #10. (**A**) Copy number profile of chromosome 19 of CLEC oligoclone #10 with transgene integration at chromosome 8p22 generated on Affymetrix Cytoscan HD Array Set. Weighted $\log_2$ signal intensity ratios and copy number state across chromosome 19 are depicted on the vertical axis, and cytobands and distance from the start of chromosome on the horizontal axis. A single locus with significant amplification (identified by at least 50 consecutive probes showing concordant change) has been circled (in red). (**B**) Characteristics of the significantly amplified genomic region. The genomic location, physical size, number of probes and copy number state across the affected region in chromosome 19 as well as the single gene residing within this region are tabulated.

2.2.2.5. Investigation of tumorigenicity and durability of FVIII expression of genome-modified CLECs in NSG mice

The tumorigenic potential of hybrid FVIII-expressing clonal CLECs with transgene integration at 8p22 locus was investigated by implantation into NOD-SCID *IL2Rγ* null (NSG) mice. The survival and engraftment of implanted cells was monitored by measuring levels of transgenic hybrid FVIII (secreted by implanted CLECs) in murine plasma using a FVIII-antigen capture ELISA assay specific for human but not murine FVIII. Untreated mice and mice implanted with CLECs that received electroporation only served as negative controls for tumor formation and plasma hybrid FVIII measurements, respectively. NSG mice implanted with a tumorigenic cell line (Hs746T) served as positive controls for tumorigenicity studies.

Hybrid FVIII was readily detected at high levels in murine plasma 3 days post-implantation (335.65 ± 19.98 mUnit/ml) (**Figure 21A**). However, a decline in plasma hybrid FVIII levels was detected on days 7 (236.9 ± 24.67 mUnits/ml) and 14 (40.17 ± 6.29 mUnits/ml) post-implantation. Although day 14 plasma FVIII levels were significantly higher ($P<0.05$) in mice implanted with FVIII-secreting CLECs (40.17 ± 6.29 mUnits/ml) compared with untreated mice (10.5 ± 2.45 mUnits/ml) or mice implanted with unmodified CLECs (8.67 ± 1.06 mUnits/ml), plasma FVIII levels were no longer different from these control mice by day 30 post-implantation. Given the highly immunodeficient nature of these mice, the likely reason for decline in FVIII levels could be a failure of implanted cells to efficiently engraft. Angiogenesis is crucial for cell survival and engraftment. Poor vascularisation (rather than development of inhibitory antibodies in immunodeficient mice) and loss of cell viability at implantation sites are potential reasons for the decline of plasma hybrid FVIII levels. Untreated mice and mice receiving CLECs that were electroporated only (without plasmid DNA) had very low levels of plasma hybrid FVIII (6 – 11 mUnits/ml), that were taken to be background levels detected of the assay method.

Although loss of viability of implanted cells could be inferred from the decline in plasma FVIII levels observed over the 30-day period following implantation, we were able to demonstrate the presence of CLECs at implantation sites by immunohistochemical staining for human vimentin. **Figure 21B** shows human vimentin staining of excised Matrigel implants from implanted mice at 1, 2 and 3 months post-implantation. We were also able to derive secondary cultures of the implanted cells from recovered Matrigel implants which were subsequently shown to be human vimentin positive by immunohistochemical staining (**Figure 21C**) and to be secreting FVIII (1304 ± 34.05 mUnits FVIII/ 10^6 cells/ 24 hr) in

174

culture. Collectively, these data suggest that a proportion of the initially implanted cells do survive and establish long-term engraftment but are not sufficiently numerous to secrete detectable levels of transgenic FVIII in murine plasma. While no tumors were observed for up to 3 months post-implantation from implantation of 3 x 10^6 unmodified or FVIII-expressing CLECs (n=20 mice), large nodular tumors were readily observed in mice implanted (n=4) with similar numbers of the tumorigenic cell line, as early as 3 weeks post-implantation (**Figure 21D**). These data suggest that oligoclonal CLECs with transgene integration at 8p22 locus which had been cultured *in vitro* for at least 2 months prior to implantation were unlikely to be tumorigenic.
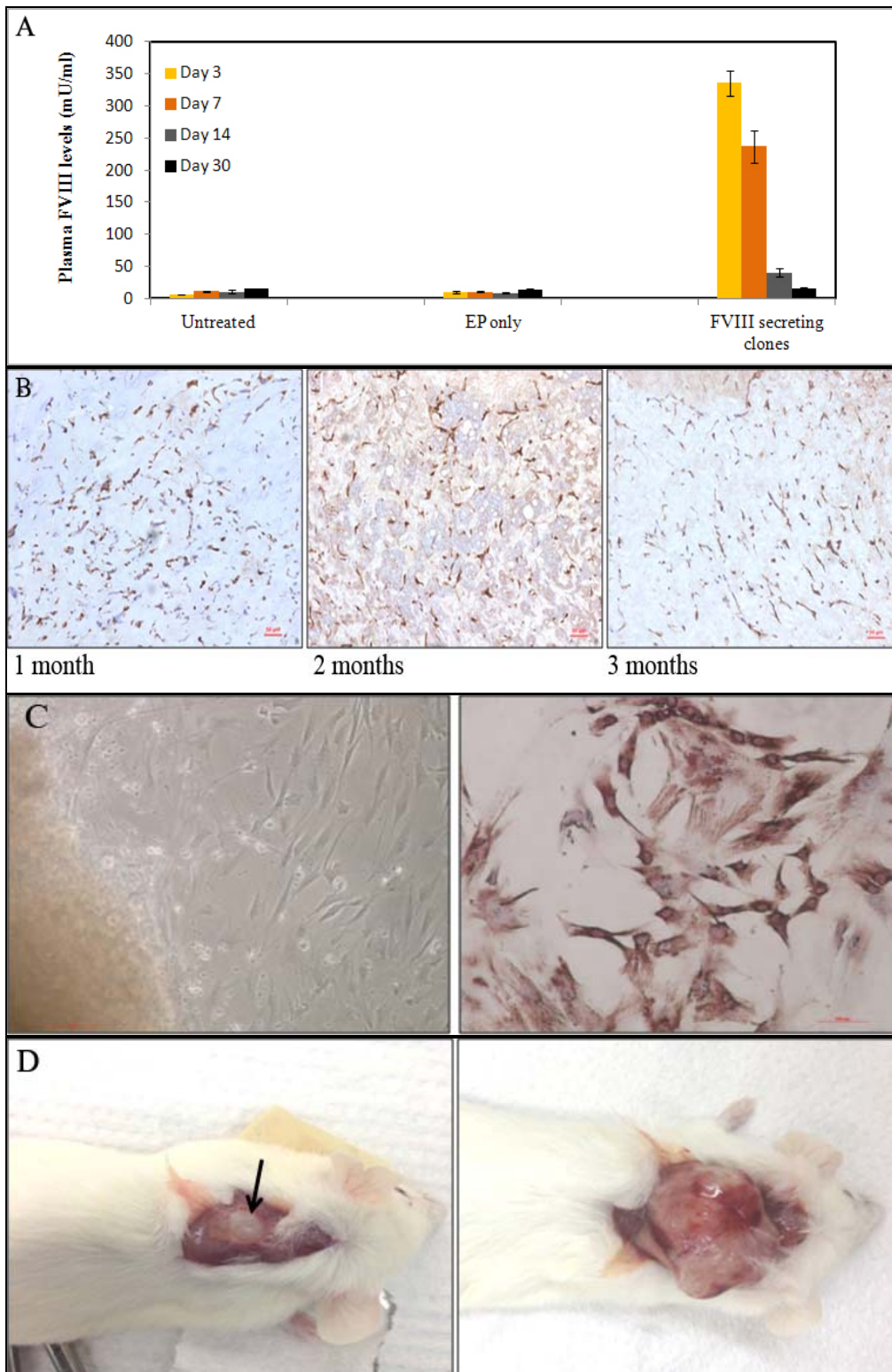
Figure 21 **Assessment of *in vivo* FVIII secretion and tumorigenicity following implantation of CLECs stably secreting transgenic hybrid FVIII in NSG mice. (A)** Plasma FVIII levels of NSG mice that were unimplanted (n=4) or implanted subcutaneously with (i) Matrigel-encapsulated CLECs that received electroporation only (EP only, n=4); or (ii) FVIII-secreting oligoclonal CLECs (n=16) with hybrid FVIII cDNA integrated at 8p22 locus (3 x 10^6 cells per animal) were measured using an ELISA based on antigen capture specific for human FVIII, on days 3, 7, 14 and 30 post-implantation. Plasma FVIII levels of mice implanted with

FVIII-secreting cells were significantly higher ($P < 0.05$) on days 3 (yellow bar), 7 (orange bar) and 14 (grey bar) compared to control mice that were either unimplanted or implanted with cells electroporated without any plasmid DNA (EP only). Data are mean ± SEM. (**B**) Immunohistochemical staining of Matrigel implants retrieved from implanted mice at the indicated time points post-implantation show evidence (brown staining) for the presence of engrafted CLECs expressing human vimentin. (**C**) (**Left**) Brightfield view of outgrowth cells from a culture of Matrigel implants retrieved 2 months post-implantation; and (**right**) immunohistochemical staining of explant cultured cells for human vimentin expression. (**D**) No tumors were observed in mice implanted with Matrigel encapsulated 3 x10^6 genome-modified CLECs at day 90 post-implantation (**left**) while nodular tumors were observed in mice implanted with equal numbers of Matrigel-encapsulated a tumorigenic cell line, Hs746T, 1 month post-implantation (**right**). Arrow points to Matrigel implant.

2.2.3. Stable transgene integration and expression in human dermal fibroblasts, bone marrow-derived and adipose-derived stromal cells

Having documented the capacity for site-specific transgene integration and stable expression of hybrid FVIII in primary human CLECs, we next investigated if other adult primary human cell types could also be modified using the phiC31 integrase to stably integrate and durably express the hybrid FVIII transgene.

Bone marrow-derived stromal cells (BMSC), adipose-derived stromal cells (ADSC) and normal dermal fibroblasts (NF123) (all primary cultures) were co-electroporated with pattB Hfer hybrid FVIII and phiC31 integrase plasmids, and subsequently selected with G418 to derive stable cell cultures. Integration junction PCR performed on genomic DNA extracted from these stable cell cultures provided evidence for site-specific transgene integration at 8p22 locus (**Figure 22A**), showing that phiC31 integrase-mediated transgene integration into the 8p22 locus was feasible in a range of adult primary human cell types.

Consistent with stable transgene integration into the genome, durable FVIII expression was observed *in vitro* in all three stable cell cultures for up to 1 month post-electroporation (duration of this experiment) (**Figure 22B**). These results suggest that the human ferritin light chain promoter was not silenced and could stably drive hybrid FVIII expression following transgene integration into genomic regions to support durable transgene expression. We are cognizant that different cell types may require different promoters for optimal activity.

Thus, the results suggest that phiC31 integrase-mediated transgene integration can induce durable transgene expression and that this may be applicable for cell- and gene-based therapies for hemophilia A delivered *via* a range of different primary adult human cell types.
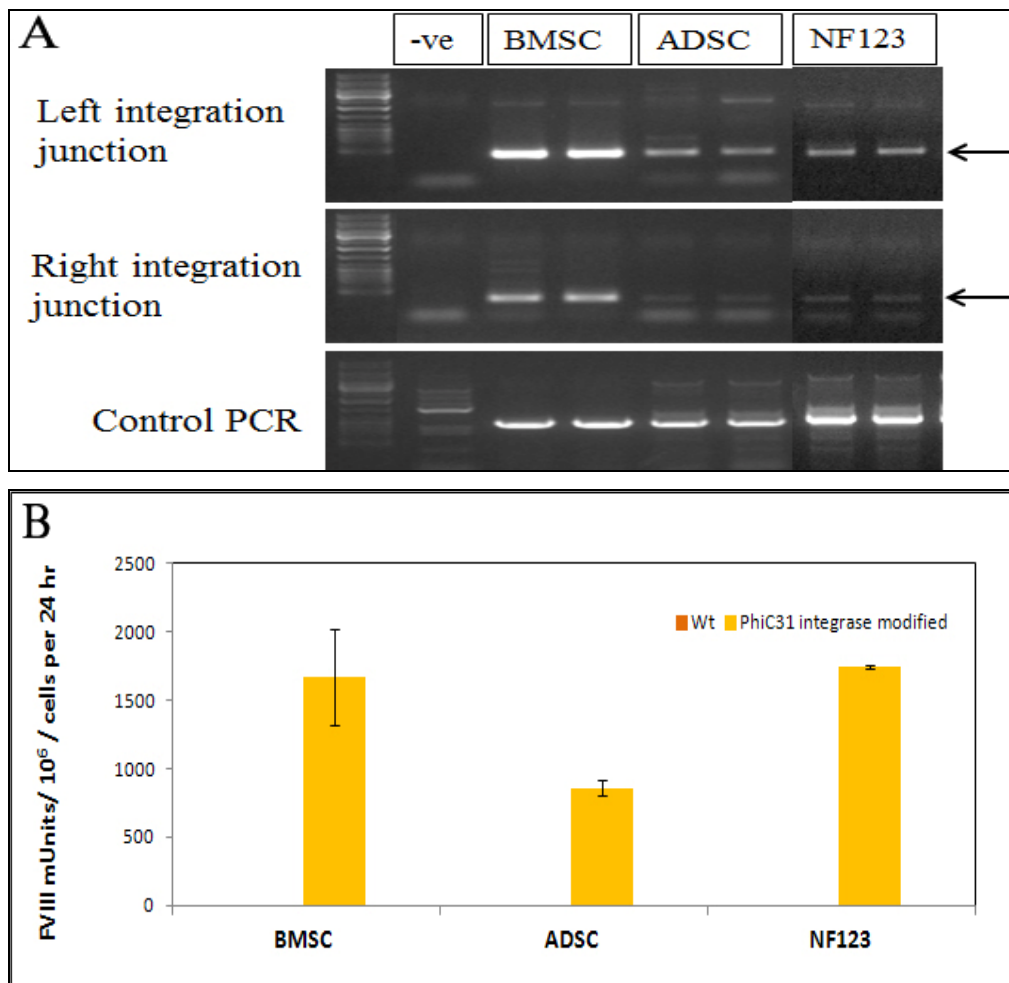
Figure 22 **Transgene integration at the 8p22 locus and durable FVIII secretion in phiC31 integrase-modified primary adult human cells.** (**A**) G418-resistant stable cells derived from primary cultures of bone marrow-derived stromal cells (BMSC), adipose-derived stromal cells (ADSC) and normal dermal fibroblasts (NF123) co-electroporated with pattB hybrid FVIII and phiC31 integrase were examined for evidence of transgene integration at 8p22 locus by left and right integration junction PCR. "-ve" refers to minus template amplification. Left and right integration junctions were amplified with vector-specific and genomic DNA-specific primers while control genomic PCR was performed with a pair of genome specific-primers amplifying a 900-bp region within chromosome 19. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Black arrows indicate the predicted integration junction PCR amplified bands. (**B**) FVIII activity in overnight conditioned media of untreated (Wt) and G418-resistant (phiC31 integrase-modified) BMSC, ADSC and NF123, determined 1 month after co-electroporation with pattB hybrid FVIII and phiC31 integrase. Data are mean ± SEM; n= 3 per group.

2.2.4. References

1. Doering CB, Healey JF, Parker ET, Barrow RT*, et al*. Identification of porcine coagulation factor VIII domains responsible for high level expression via enhanced secretion. J Biol Chem. (2004) 279:6546-6552.

2. Doering CB, Denning G, Dooriss K, Gangadharan B*, et al*. Directed engineering of a high-expression chimeric transgene as a strategy for gene therapy of hemophilia A. Mol Ther. (2009) 17:1145-1154.

3. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B*, et al*. Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol. (2006) 357:28-48.

4. Liao YC, Lo SH. Deleted in liver cancer-1 (DLC-1): a tumor suppressor not just for liver. Int J Biochem Cell Biol. (2008) 40:843-847.

5. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. (2001) 25:402-408.

6. Jiang Z, Huang Y, Sharfstein ST. Regulation of recombinant monoclonal antibody production in chinese hamster ovary cells: a comparative study of gene copy number, mRNA level, and protein expression. Biotechnol Prog. (2006) 22:313-318.

7. Wang GP, Garrigue A, Ciuffi A, Ronen K*, et al*. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res. (2008) 36:e49.

2.3. AAVS1 ZFN cell modification

Data detailed in the preceding sections showed that phiC31 integrase could stably modify CLECs for durable transgene expression and raised the possibility of developing FVIII-secreting CLECs as bioimplants for hemophilia therapy. Although most of our genotoxicity data suggested minimal risk for adverse outcomes following phiC31 integrase-mediated transgenesis, the known capacity of this system for integrations into multiple genomic regions and our observation of low frequency chromosomal translocations prompted us to evaluate ZFNs as a more precise and thus safer alternative for site-directed transgenesis into a potentially safe genomic harbor, the AAVS1 locus[1].

2.3.1. AAVS1 ZFN-mediated homologous recombination in K562 cells

Initial optimization and evaluation of AAVS1 ZFNs to integrate donor DNA of varying sizes was evaluated in a human chronic myelogenous leukemia cell line, K562, which other investigators have shown to be permissive for ZFN-mediated transgene integrations at high efficiencies[2]. Our initial studies were performed with commercially purchased AAVS1 ZFN mRNA (Sigma-Aldrich).

2.3.1.1. Integration of 50-bp donor DNA into AAVS1 locus

The first evidence of site-specific integration at the AAVS1 locus of K562 cells was the demonstration of integration of a 50-bp donor DNA fragment following co-transfection with AAVS1 ZFN mRNA and pZDonor plasmid. The transfection efficiency as assessed by flow cytometry following nucleofection of an EGFP reporter gene was 49.54 ± 0.26 % (Data are mean $\pm$ SEM, n=3).

Junction PCR analysis using a vector-specific primer and AAVS1 genomic DNA specific primer provided evidence for site-directed integration of the 50-bp donor DNA in K562 cells only when co-electroporated with pZDonor and AAVS1 ZFNs (Donor + AAVS1 D4) and not when only pZDonor (Donor only) was electroporated. Treated cells (Donor + AAVS1 ZFN) analyzed on day 4 and day 16 post-treatment were positive for site-specific integration, demonstrating stable modification of cells (**Figure 23, junction PCR**). Control genomic PCR amplifying a common AAVS1 locus present in all cell populations (Donor only and Donor + AAVS1 ZFN), verified the integrity of genomic DNA template and served as a positive PCR control.

Site-specific integration of 50-bp donor DNA at the AAVS1 locus was confirmed by RFLP assay. The two faster migrating bands (**Figure 23, RFLP assay,**

Donor + AAVS1 ZFN D4), restriction enzyme-digested PCR products, provided evidence for site-specific integration of a 50-bp donor DNA fragment carrying a novel restriction enzyme site not present at the endogenous AAVS1 locus. Genomic DNA extracted from K562 cells electroporated with donor DNA only were negative for RFLP PCR products (**Figure 23, RFLP assay, Donor only**), indicating absence of site-specific integration of donor DNA without co-electroporation of ZFN. Site-specific double-stranded DNA breaks induced by AAVS1 ZFNs are required to greatly enhance homologous recombination (HR) and site-specific integration of donor DNA. Densitometric measurements of cleaved, restriction digested RFLP PCR products (two faster migrating bands) expressed as a percentage of total DNA (cleaved products and uncleaved DNA) gave an estimate of the percentage of K562 cells that attained site-specific integration. Such an analysis revealed approximately 52.64 ± 0.85 % of treated K562 cells (data are mean ± SEM, n=2) attaining site-specific donor DNA integration when co-transfected with donor DNA and AAVS1 ZFN.
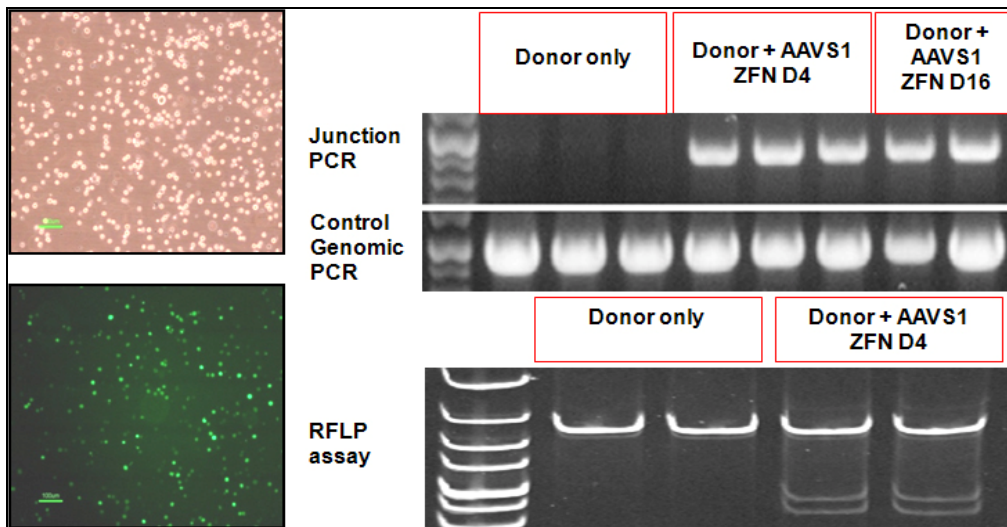
Figure 23 ZFN-mediated site-specific integration of 50-bp donor DNA into AAVS1 locus in K562 cells. K562 cells were co-electroporated with pEGFP-C1 and pZDonor plasmid DNA with or without AAVS1 ZFN mRNA. (**Left**): Brightfield and fluorescent images of transfected K562 cells. PCR spanning the integration junction (**top**) was performed (in duplicates or triplicates) on genomic DNA from cells treated with donor DNA only (4 days post- treatment) or with donor DNA and AAVS1 ZFN (4 and 16 days post-treatment), with genome specific (AAVS1 R) and vector specific primers (MCS F)**.** Control genomic PCR amplified an adjacent 900-bp region of the AAVS1 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. (**Bottom**): RFLP assay was performed (in duplicate) by digesting genome amplified PCR products (amplified with genome-specific primers spanning the integration site) with *Hin*d III restriction enzyme followed by electrophoresis in a 5% polyacrylamide gel. Site-specific integration of 50-bp DNA results in introduction of a *Hin*d III recognition site which is evident by the appearance of two cleaved, faster migrating DNA bands. Densitometric measurements (performed with Quantity One software) of digested bands (modified) expressed as a percentage of undigested (unmodified) and digested bands gave the estimated site-specific integration frequency.

2.3.1.2.   Optimization of ZFN activity and HR frequency

To optimize ZFN activity and improve HR frequency, we evaluated the effects of mild hypothermic incubation and different AAVS1 ZFN variants on the efficiency of ZFN-mediated site-specific transgene integration.

2.3.1.2.1.   JPCR evaluation on the effects of AAVS1 ZFN variants and mild hypothermia

Commercially purchased AAVS1 ZFN mRNAs which were costly and of limited quantity were compared against AAVS1 ZFNs cloned as plasmids and expressed from the CMV promoter in this project. ZFNs were mutated to function as obligate heterodimers[3] (OH ZFNs), further modified to enhance *FokI* cleavage activity[4] (Sharkey ZFNs) and modified with a third mutation[5] to further improve nuclease activity (termed enhanced Sharkey ZFNs). We also compared the effect of expressing left and right ZFNs from a single vector compared to expressing each ZFN from its own plasmid. Site-specific integration of a 50-bp donor DNA from pZDonor was evaluated by junctional PCR analysis and RFLP assay.

Junctional PCR band intensities normalized to corresponding control genomic PCR band intensities and expressed as a percentage of AAVS1 ZFN mRNA treated samples showed increased site-specific integrations mediated by enhanced Sharkey ZFNs (84%) and Sharkey ZFNs (52%) compared to obligate heterodimeric ZFNs (36%) (**Figure 24**). Given the lower probability of 2 single vector ZFNs being transfected into the same cell by electroporation compared to a single vector expressing both ZFNs, it was not surprising to observe an almost two-fold increase in site-specific integration when Sharkey ZFNs were delivered as a dual expression cassette (Sharkey Dual ZFN, 94%) compared to 2 single vector ZFN plasmids (Sharkey 2 single ZFNs, 52%). Incubation of cells under mild hypothermic conditions[6] (OH 2 single ZFNs 30°C) resulted in an almost 2-fold increase in site-specific integration (80% vs 43%) compared to incubation at 37°C (OH 2 single ZFNs 37°C), as reported previously[6].
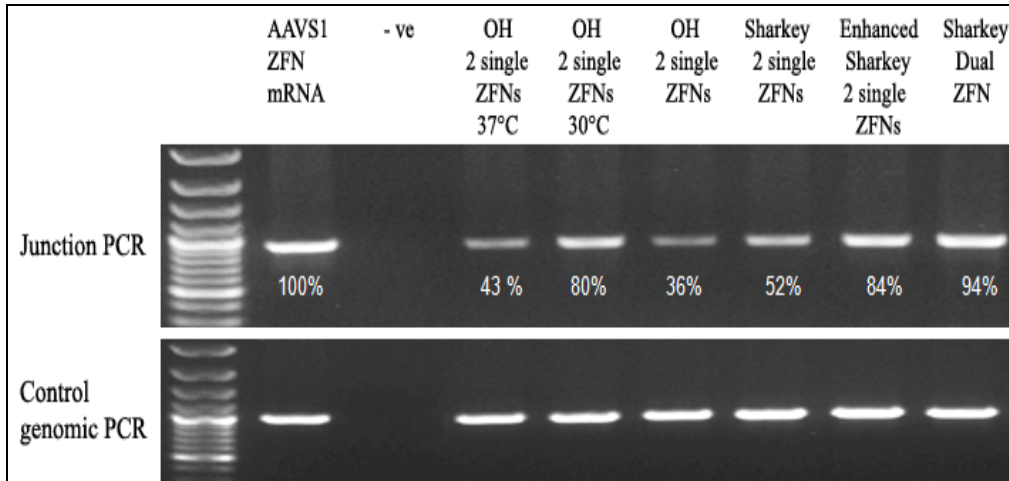
Figure 24   Optimization of site-specific integration with different AAVS1 ZFN plasmids. K562 cells co-transfected with 10 μg of pZDonor and different combinations of AAVS1 ZFNs (5 μg in total) were examined by junction PCR 4 days post-treatment (incubated at 37°C unless otherwise indicated) for site-specific integration of a 50-bp donor DNA. AAVS1 ZFNs were delivered as mRNA (Sigma) or as obligate heterodimeric (OH) ZFNs, Sharkey ZFNs and enhanced Sharkey ZFNs delivered from 2 single vector plasmids or Sharkey ZFNs delivered as a dual expression vector. Control genomic PCR amplified a 900-bp region of the AAVS1 locus and minus template PCR (-ve) served as a negative PCR control. Amplified products were electrophoresed on 1% agarose gels, imaged using BioRad®Gel Doc 2000 transilluminator and analyzed by densitometric measurements using QuantityOne software. Densitometric measurements of junctional PCR bands were normalized to their corresponding control genomic PCR bands. For each treatment, the normalized band intensities was expressed as a percentage of the AAVS1 ZFN mRNA sample.

2.3.1.2.2.   RFLP evaluation on the effects of AAVS1 ZFN variants and mild hypothermia

Results of junctional PCR analyses were confirmed by performing RFLP assays on K562 cells treated with either AAVS1 ZFN mRNA or ZFN variant plasmids delivered from a dual expression vector. The effect of incubation in mild hypothermic conditions was also investigated. **Figure 25** shows RFLP products resolved on PAGE gels and quantified by densitometry to calculate percent donor DNA integration shown in panel B.

Densitometric measurements of RFLP assay products showed that at 37°C, K562 cells treated with enhanced Sharkey ZFNs (25.29 ± 0.58%) and Sharkey ZFNs (19.89 ± 0.58%) achieved significantly greater integration of donor DNA ($P<0.05$) compared to OH ZFN treated samples (12.22 ± 1.83%). For all 3 ZFN variants, significantly improved donor DNA integration ($P<0.05$) was achieved when cells were transiently (2 days) incubated at 30°C [OH 30°C: 26.97 ± 1.35%; Sharkey 30°C: 30.73 ± 1.7%; enhanced Sharkey 30°C: 32.39 ± 0.53%] compared with 37°C [OH 37°C: 12.22 ± 1.83%; Sharkey 37°C: 19.89 ± 0.58%; enhanced Sharkey 37°C:25.29 ± 0.58%]. The effect of mild hypothermic incubation was more pronounced for cells treated with OH ZFNs (2-fold) compared with Sharkey (1.5-fold) or enhanced Sharkey ZFNs (1.3-fold), suggesting saturation of ZFN cleavage activity under conditions of mild hypothermia plus enhanced Sharkey ZFN treatment. Furthermore, at 30°C enhanced Sharkey ZFN resulted in significantly greater site-specific integrations compared to only OH ZFN ($P = 0.02$) but not Sharkey ZFN ($P>0.05$) treatment. Overall, optimal conditions for site-specific integration of donor DNA were treatments with either Sharkey ZFNs or enhanced Sharkey ZFNs combined with mild hypothermia.

Direct comparison of AAVS1 ZFN variant plasmid DNAs to commercially purchased AAVS1 ZFN mRNA was not possible because of significant difference in transfection efficiency. Further analysis by normalizing percent donor integration to a common transfection efficiency of 60% revealed that at 37°C, only enhanced Sharkey ZFN treatment resulted in significantly ($P = 0.0007$) increased donor integration (1.3-fold) compared to AAVS1 ZFN mRNA treatment. It was not possible to determine whether any of our AAVS1 ZFN DNA variants were superior to commercially purchased mRNA because the eventual absolute quantity of ZFN transcripts (mRNA) generated from delivered ZFN plasmid DNA constructs was unknown. Thus, the difference in frequency of donor DNA integration between the mRNA and DNA versions of ZFNs may be due to differences in the final mRNA quantity and eventual

protein amount expressed in treated cells. Nevertheless, due to cost constraints of using AAVS1 ZFN mRNA, all further experiments were performed using plasmid constructs encoding enhanced Sharkey ZFNs combined with mild hypothermia.
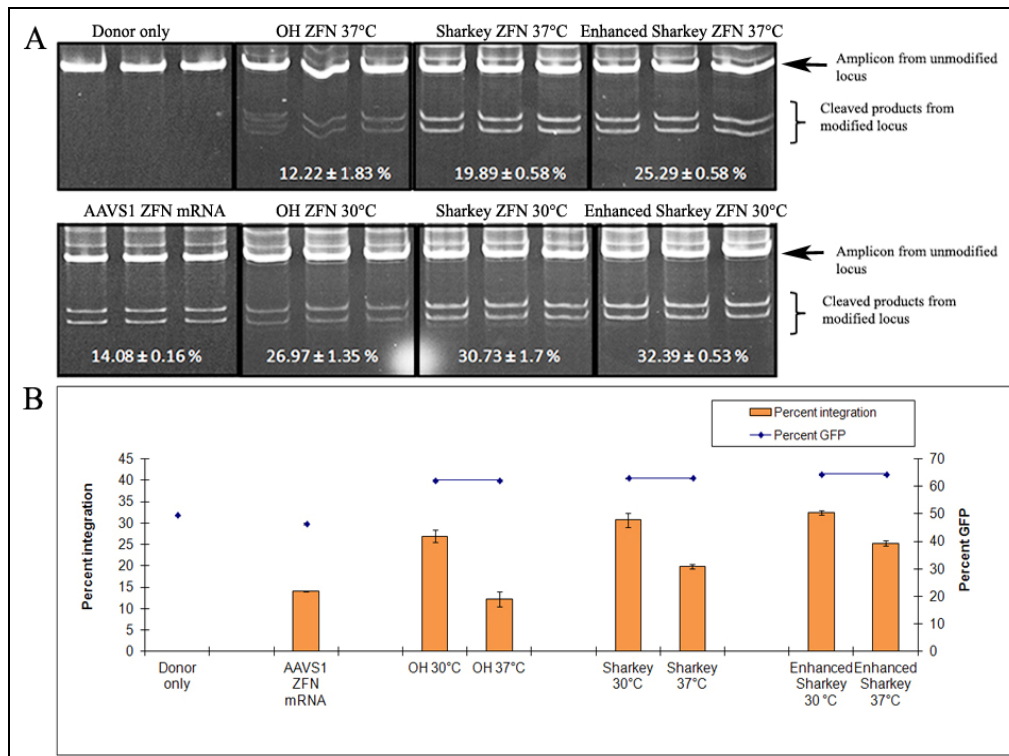
Figure 25   Effect of ZFN variants and mild hypothermia on site-specific donor DNA integration in K562 cells. (**A**) Genomic DNA extracted from K562 cells co-electroporated with plasmid encoding EGFP, pZDonor and the following AAVS1 ZFN variants; AAVS1 ZFN mRNA, OH (obligate heterodimer), Sharkey (OH modified according to Guo J., *et al*)[4] or enhanced Sharkey (Sharkey ZFN variant further modified according to Doyon Y., *et al*)[5] and cultured at either 37°C or 30°C, were amplified with primers spanning the AAVS1 locus, digested with *Hin*dIII, electrophoretically resolved in 5% polyacrylamide gels, stained with ethidium bromide and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. The values displayed at the bottom of the gels are mean densitometric measurements (of triplicates) of the modified locus expressed as a percentage of the unmodified and modified locus combined (Data are mean ± SEM; *n* = 3). (**B**) Percentage of K562 cells (as determined by densitometric measurements of RFLP assay products) attaining AAVS1-specific integration of a 50-bp donor DNA (left axis) and the percentage of GFP-positive cells (right axis) as analyzed by FACS are shown. Data are mean ± SEM (n=3). Percent integration was significantly greater ($P<0.05$) for cells incubated at 30°C compared to 37°C for all treatment groups. At 37°C, significantly greater integration was achieved with enhanced Sharkey ($P = 0.0024$) and Sharkey ($P = 0.03$) compared to OH. At 30°C, enhanced Sharkey was significantly different ($P = 0.02$) from OH but not from Sharkey.

2.3.1.3. Integration of 4-kb EGFP transgene cassette

Having optimized conditions for efficient ZFN expression we next attempted integration of a larger donor DNA fragment, a 4 kb cassette consisting of EGFP and neomycin resistance cDNAs (pZDonor EGFP), into the AAVS1 locus of K562 cells.

Left and right integration junction PCR analysis on genomic DNA extracted from unselected or G418-selected K562 cells co-electroporated with pZDonor-EGFP and enhanced Sharkey ZFN, provided evidence for integration of the 4-kb donor DNA at the AAVS1 locus (**Figure 26**). Increased intensities of left and right junctional PCR amplicons in G418-selected cell samples (K562 + AAVS1 ZFN + selection) indicated enrichment of K562 cells stably integrated with the transgene at the AAVS1 locus.

Sequencing and analysis of integration junction PCR amplicons revealed precise integration of vector sequences into the AAVS1 genomic locus recognized and cleaved by AAVS1 ZFNs (AAVS1 recognition sequences underlined in red and integrated vector sequences in blue).
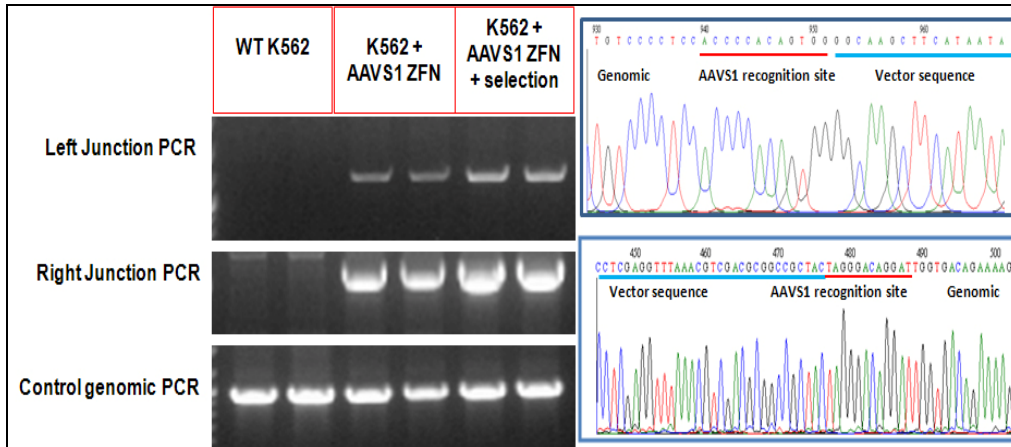
Figure 26 **ZFN-mediated site-specific integration of a 4-kb DNA cassette encoding EGFP donor DNA into AAVS1 locus in K562 cells.** K562 cells were co-electroporated with pZDonor EGFP plasmid DNA (4-kb EGFP cassette) and a bicistronic plasmid encoding both left and right AAVS1 ZFNs (enhanced Sharkey). (**Left**) Genomic DNA from untreated cells (WT K562) and K562 cells treated with donor DNA and AAVS1 ZFN with or without G418 selection were investigated for site-specific integrations of the vector DNA by integration junction PCR. Left and right integration junctions were amplified with vector specific and genomic DNA specific primers while control genomic PCR was performed with a pair of genome-specific primers amplifying a 900-bp region adjacent to the ZFN-targeted site in the AAVS1 locus (all done in duplicates). (**Right**) DNA sequence chromatograms of (**top**) left and (**bottom**) right junctional PCR amplicons. Vector sequences are underlined in blue while AAVS1 ZFN recognition half-sites are underlined in red.

2.3.1.4.   Integration of 9-kb hybrid FVIII transgene cassette

Our goal with ZFN-mediated transgenesis was to integrate a 9-kb DNA fragment encoding a B-domain deleted hybrid-FVIII and neomycin resistance cDNA at the AAVS1 locus.

We were able to achieve AAVS1 site-specific integration of the 9-kb donor DNA fragment in G418-selected K562 samples co-transfected with pZDonor hybrid FVIII and enhanced Sharkey ZFNs (with mild hypothermia incubation). Left and right junction PCR screening (**Figure 27**) showed positive amplifications (indicative of vector integration) in cells co-transfected with ZFNs (K562 + Donor + AAVS1) but were absent in samples transfected with donor DNA alone (K562 + Donor only), demonstrating the requirement for ZFN activity in mediating site-specific integration.

Sequencing of junction PCR amplicons showed no deletions to either the genomic DNA or integrated donor at the AAVS1 locus, thus verifying the precise nature of transgene integration.
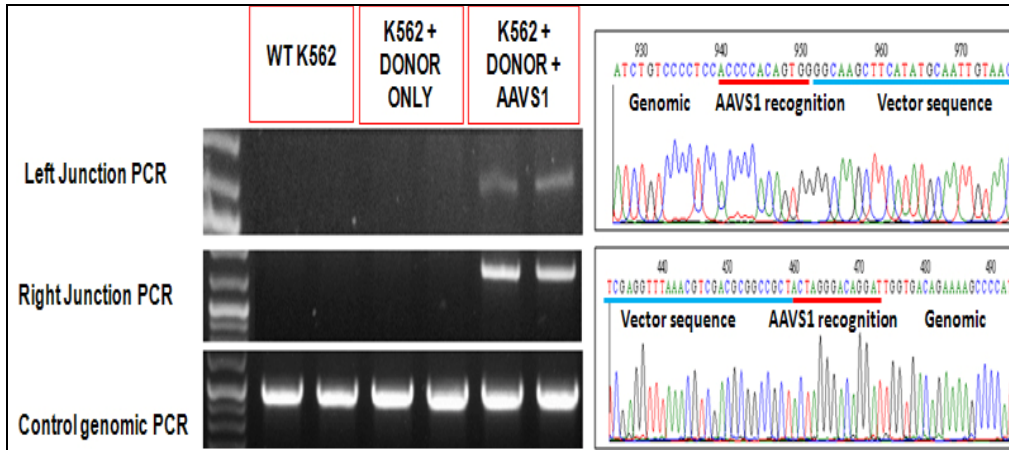
Figure 27 **ZFN-mediated site-specific integration of a 9-kb DNA cassette encoding hybrid FVIII donor DNA into AAVS1 locus in K562 cells.** K562 cells were either electroporated with donor plasmid DNA (9-kb hybrid FVIII cassette) only or co-electroporated with a bicistronic plasmid encoding both left and right AAVS1 ZFNs. (**Left**) Genomic DNA from untreated cells (WT K562) and K562 cells treated with donor DNA only or with AAVS1 ZFNs were investigated for site-specific integrations of the vector DNA by integration junction PCR. Left and right integration junctions were amplified with vector-specific and genome-specific primers, while control genomic PCR was performed with a pair of genome-specific primers amplifying a 900-bp region adjacent to the ZFN-targeted site in the AAVS1 locus (all done in duplicates). (**Right**) DNA sequence chromatograms of left (**top**) and right (**bottom**) junctional PCR amplicons. Vector sequences are underlined in blue while AAVS1 ZFN recognition half-sites are underlined in red.

2.3.1.5.   Efficiency of site-specific integration

The efficiency of ZFN in mediating site-specific integration of a 9-kb donor DNA was estimated by performing direct integration junction PCR analysis on random clonal populations of K562 cells from a bulk population of unselected or G418-selected K562 cells treated with pZDonor Hybrid FVIII and enhanced Sharkey AAVS1 ZFNs. Single cells sorted into individual wells of a 96-well plate were expanded in culture, lysed *in situ* and used as crude genomic template for PCR amplification using Phusion® Human Specimen Direct PCR kit. The efficiency of site-specific integration was given by the percentage of total analyzed clonal cells positive for site-specific integrations (identified by right integration junction PCR). Only control PCR positive samples were considered valid.

Direct PCR for site-specific integration was performed on 40 unselected and 79 G418-resistant clonal K562 cells. Our analyses identified 4 of 40 (10%) of unselected K562 clones (**Figure 28A**) and 22 of 79 (27%) of G418-resistant clones (**Figure 28B**) as positive for site-specific integrations of the 9-kb donor DNA. Thus, G418 selection resulted in an almost 3-fold enrichment for cells with site-specific integration. However, integration junction PCR could not be used to distinguish monoallelic from biallelic transgene integrations. Moreover, K562 cells are trisomic for chromosome 19. Lack of complete G418 selection of suspension cultures of K562 cells or selection of cells with random transgene integrations could account for 63% of G418-selected clones being negative for site-specific integration.

We further analyzed clones that were positive for the left integration junction by screening for the right integration junction (**Figure 28C**). Intriguingly, our results showed that of 22 such clones, only 13 had both integration junctions. This raised the possibility of deletion of either genomic or vector DNA that resulted in the loss of primer binding sites at integration junctions. Of more serious concern was the possibility that some clones may have sustained only partial integration of the transgene cassette which would render them useless for transgene expression. This possibility of partial transgene integration of the transgene was later evaluated in CLECs by segmental PCR to determine if all the essential components of the donor vector for FVIII transgene expression were present in clonal cells that had either a single or both integration junctions (**section 2.3.2.8**).
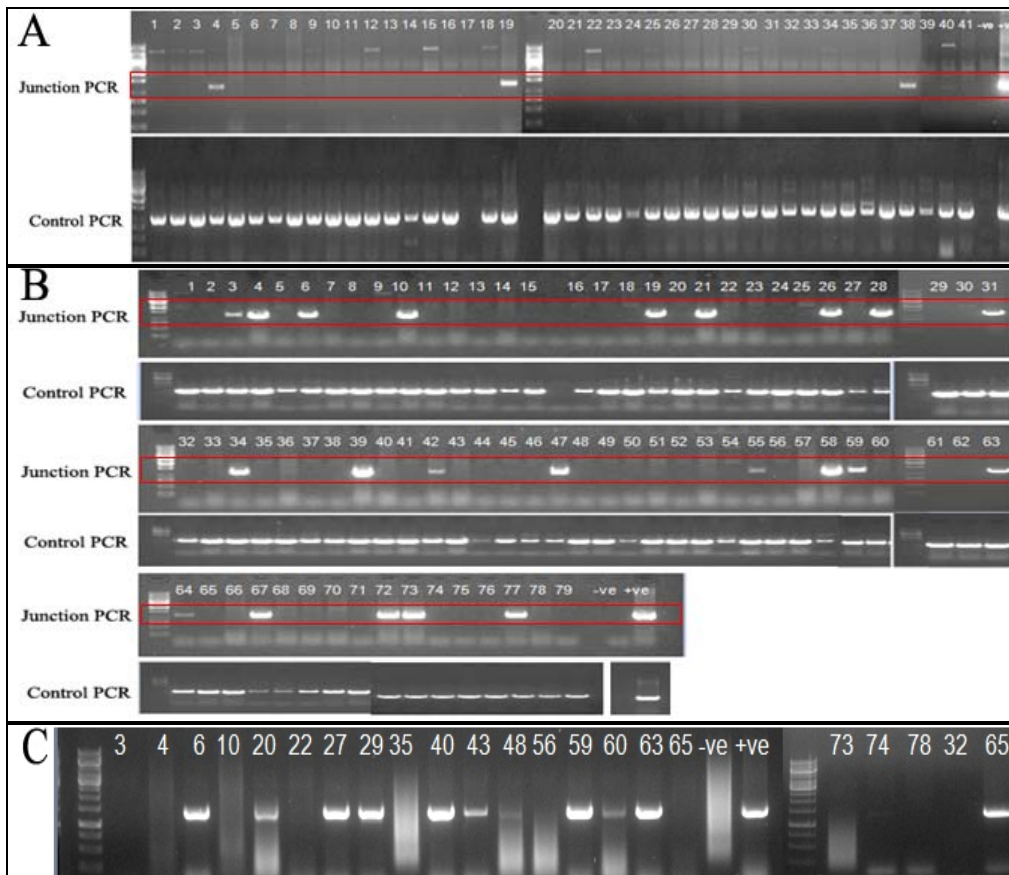
Figure 28   Investigation of a clonal population of K562 cell-line for site-specific integration of hybrid FVIII DNA cassette at the AAVS1 locus. Left integration junction PCR and control PCR (amplifying a -kb genomic region at AAVS1 locus) were performed on FACS-sorted clonal populations of either (**A**) unselected or (**B**) G418-selected K562 cells integrated with pZDonor hybrid FVIII plasmid. (**C**) Clones from G418 selected K562 which were identified as positive for left integration junctions were analyzed for right junction PCR. *In situ* lysed cells were starting materials for PCR amplification using Phusion[®] Human Specimen Direct PCR kit. Amplification products were electrophoresed on 1% agarose gels and imaged using BioRad[®]Gel Doc 2000 transilluminator. Positive (+ve) and negative (-ve) controls for junction PCR were genomic DNA sample from a pooled population of K562 with pZDonor Hybrid FVIII integration and minus template, respectively. For each gel, red rectangles demarcate the predicted position (i.e. size) of positively amplified integration junction amplicons.

2.3.2. Evaluation of AAVS1 ZFN-mediated HR in primary human cells

2.3.2.1. Optimization of gene transfer to CLECs

High efficiency gene transfer is essential for optimal ZFN-mediated transgenesis. Given the modest gene transfer achieved with electroporating CLECs using the BTX® electroporation system (phiC31 integrase study), we explored electro-gene transfer of CLECS using the Amaxa® Nucleofector™ I or Nucleofector ™ 4D devices.

Figure 29A summarizes data comparing the Nucleofector™ I device and BTX® electroporation system based on percent GFP-positive cells and percent viable cells achieved. Compared with initially optimized settings using the BTX® electroporation system, significantly improved gene transfer (percent GFP, $P = 0.0003$) and cell survival (percent viability, $P < 0.0001$) were achieved using the Nucleofector™ I device, with program T-23 (percent GFP: $47.69 \pm 1.96\%$; percent viability: $51 \pm 0.67\%$; data are mean $\pm$ SEM, n=3) giving the best results. Comparatively, electroporation using optimized BTX® electroporator settings resulted in only $29.94 \pm 0.56\%$ of GFP-positive cells with $24 \pm 2.31\%$ of all treated cells being viable (Data are mean $\pm$ SEM, n=3). Thus, subsequent gene transfer studies were done using Nucleofector® I device and program T23.

During the course of this study, technological improvements were made to Nucleofector™ devices which led us to evaluate the most recent gene transfer equipment from Amaxa®, i.e. Nucleofector™ 4D device. **Figure 29B** summarizes gene transfer and viability data of CLECs electroporated with a Nucleofector ™ 4D device using several programs . Compared with results achieved with T23 program of the Nucleofector ™ I device, electroporation of CLECs using Nucleofector ™ 4D device and program CM113 resulted in an overall improved gene transfer efficiency, with $58.37 \pm 0.23\%$ of treated cells being GFP-positive ($P = 0.0025$) and $62.16 \pm 1.88\%$ of all cells being viable (Data are mean $\pm$ SEM, n=3)($P = 0.055$). Thereafter, CLECs were electroporated using Nucleofector ™ 4D and program CM113.
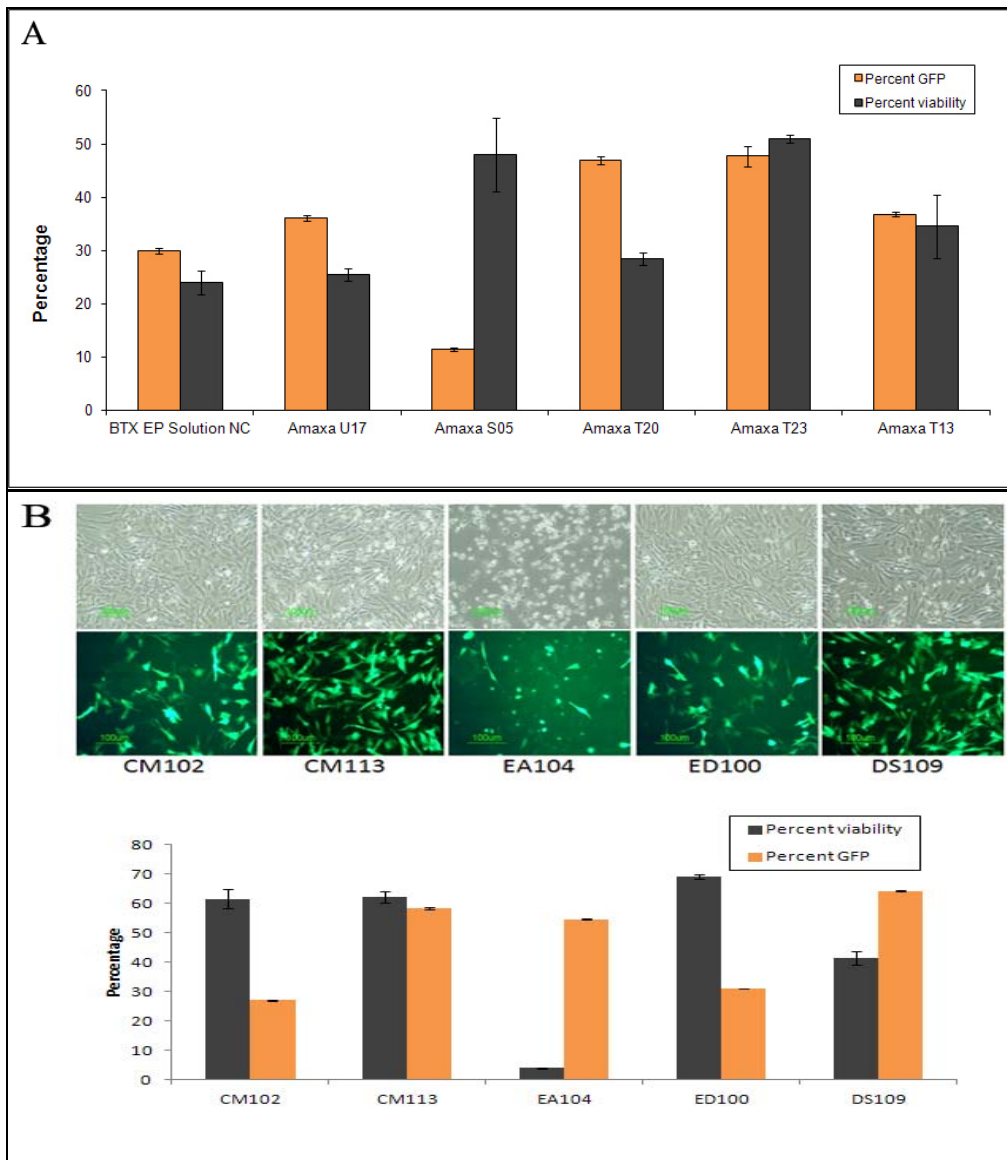
Figure 29   Optimization of electroporation with Amaxa™ Nucleofector I and 4D programs. CLECs electroporated with 10 µg of EGFP-C1 reporter plasmid, using solution NC and a single 30 ms pulse delivered at 600 V/cm with a BTX® electroporator, or using Amaxa® Basic Nucleofector™ kit for primary mammalian epithelial cells and pulses delivered by the indicated programs (S-05, T-23, U-17, T-13, T-20) with (**A**) Amaxa®Nucleofector™ I device; or (**B**) Nucleofector™ 4D device using Nucleofector™ Primary cell solution P1 and programs CM102, CM113, EA104, ED100, DS109. Electroporated CLECs, were analyzed 1 day post-electroporation for GFP expression by flow cytometry (percent GFP, yellow bar) and for viability (percent viability, grey bar) by trypan blue exclusion cell counts. Data are mean ± SEM (n=3). Representative brightfield and fluorescence images taken 1 day post-electroporation are shown (Original magnification x100).

2.3.2.2. Evaluation of CLECs for *PPP1R12C* transcript expression

An open chromatin configuration may be an essential factor for ZFNs to successfully target, bind and induce double-stranded DNA break at a specific genomic locus. Previously, we were able to demonstrate efficient ZFN-induced site-specific integration of donor DNA of varying sizes into the AAVS1 locus of K562 cells. We investigated if high expression of the endogenous *PPP1R12C* gene at the AAVS1 locus, reflective of an open chromatin configuration, correlates with efficient site-specific genomic cleavage and integration into the AAVS1 site. *PPP1R12C* transcript levels (exons 4-6) in K562 cells, primary human dermal fibroblasts, CLECs and a human embryonic stem cell line (HUES) were evaluated by quantitative RT-PCR (**Figure 30**).

Our data indicated approximately 2-fold lower *PPP1R12C* transcript levels in HUES and CLECs and 2-fold higher levels in dermal fibroblasts compared to K562 cells. Concordant with relatively lower *PPP1R12C* transcript levels, site-specific integration of a 50-bp donor DNA (as determined by RFLP assay) was at least 10-fold lower in CLECs compared to K562 cells (CLECs: 3.39 ± 0.1 %; K562: 32.39 ± 0.53%) (**Figures 25 and 34**). Dermal fibroblasts showed a higher frequency of genome cleavage than CLECs by *Cel*-1 nuclease assay (CLECs: 43.01 ± 1.93%; Fibroblast: 64.97 ± 3.48)(**Figures 33 and 54**), that was also associated with higher levels of *PPP1R12C* expression in fibroblasts. HUES cells were not tested for site-specific integration. Although our results suggest that transcriptional activity promotes efficient site-specific genome cleavage and integration, this should be verified by evaluating a wider range of different cell types. Moreover, it is likely that factors other than an open chromatin configuration may also be highly influential in determining the efficiency of site-specific integration of donor DNA.
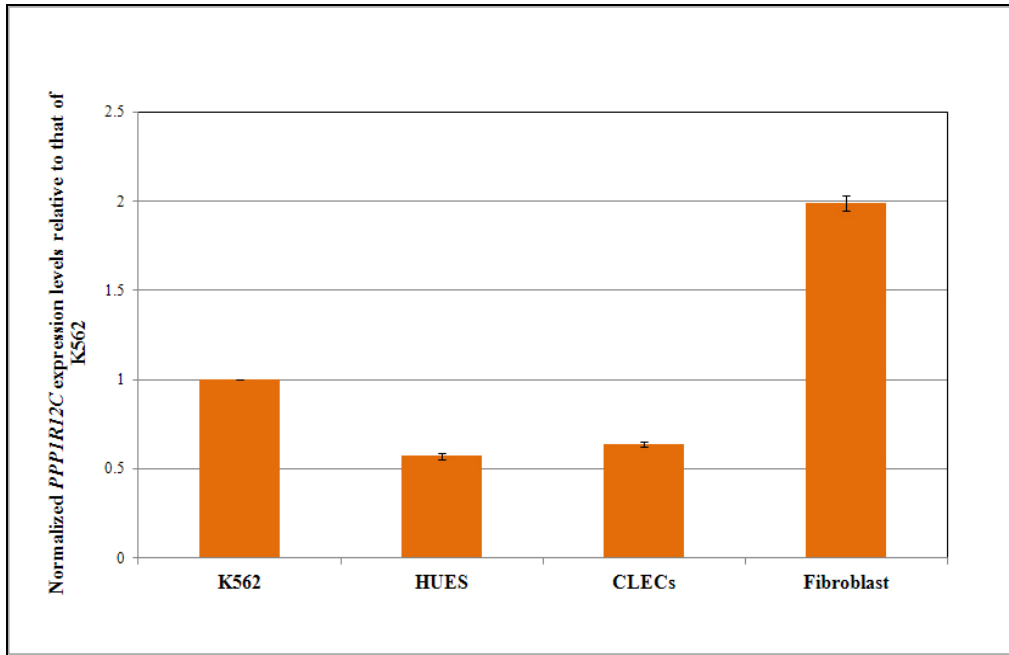
Figure 30   Quantitative RT-PCR analysis of *PPP1R12C* transcript levels in various cell types. Quantitative RT-PCR was performed on K562 cells, a human ES cell line (HUES), CLECs and primary human dermal fibroblasts to quantify *PPP1R12C* transcript levels (exon 4 -6) and *GAPDH* expression. *GAPDH*-normalized *PPP1R12C* expression levels (as determined by the $2^{-\Delta\Delta Ct}$ method) are shown relative to that of K562 cells. Data are mean ± SEM, n=2 experiments per group and 4 replicates per sample.

2.3.2.3. Evidence of ZFN expression and activity in CLECs

Before investigating ZFN-mediated transgenesis in CLECs, we first evaluated the capacity for CLECs to express the ZFNs (mRNA and protein) following transient transfection with plasmids encoding either the left or right AAVS1 ZFNs. We further confirmed the site-specific endonuclease activity of AAVS1 ZFNs by evaluating the target locus for evidence of site-specific genomic cleavage and repair (*Cel*-1 mismatch nuclease assay).

2.3.2.3.1.   RT-PCR analysis of ZFN transcripts

Reverse transcription-PCR of total RNA from treated CLECs using primers specific to left and right AAVS1 ZFN, showed evidence for expression of left and right ZFN mRNA transcripts (**Figure 31**). Densitometric measurements of RT-PCR bands were normalized to α-actin levels and expressed as a percentage of levels detected at 8 hours. Expression of both ZFN transcripts was detected from 8 to 144 hours post-electroporation, with levels decreasing to approximately 50% of initial levels (taken as 8 hours) at 144 hours post-electroporation. Greatest expression of both left and right ZFNs was observed between 8-24 hours post-electroporation.
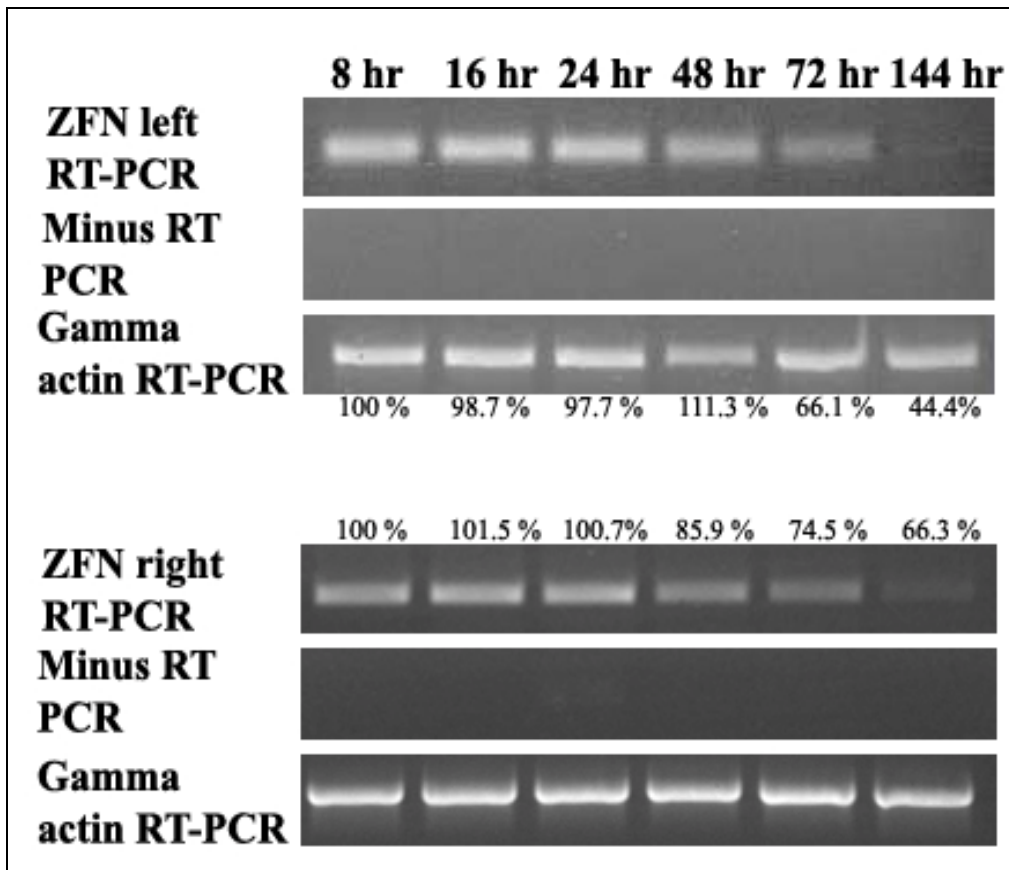
Figure 31   ZFN mRNA transcripts in CLECs following electroporation of plasmids encoding individual ZFNs. Total RNA, isolated from transfected CLECs at the indicated time points post-electroporation with plasmid DNA encoding either (**top**) left ZFN or (**bottom**) right ZFN were analyzed by RT-PCR with primers specific to the (**top**) left ZFN, (**bottom**) right ZFN or gamma-actin. The control reaction omitted reverse transcription (Minus RT PCR). Amplification products were electrophoresed on 1% agarose gels, imaged using BioRad®Gel Doc 2000 transilluminator and quantified using QuantityOne software. Densitometric measurements of ZFN transcript bands were normalized to their respective actin levels and expressed as a percentage of ZFN mRNA levels observed at 8 hours (indicated in each gel).

2.3.2.3.2.   Immunoblot identification of ZFN protein

Immunoblot detection using anti-FLAG antibodies confirmed the expression of ZFN proteins in CLECs transfected with plasmid DNA encoding AAVS1 ZFN fused with a FLAG tag but not in untransfected CLECs (WT) (**Figure 32**).

Incubation of cells under mild hypothermic conditions (30°C for 3 days or 37°C for 1 followed by 30°C for 2 days) resulted in higher levels of ZFN protein expression in transfected CLECs compared with incubation at 37°C for 3 days. Control immunoblot for β-actin expression in all samples demonstrated equal loading of total protein amongst the groups of cells analyzed. These results were consistent with the hypothesis that mild hypothermic conditions[6] may reduce the turnover rate (proteolysis) of proteins thereby resulting in higher ZFN levels under these conditions compared to conventional conditions.
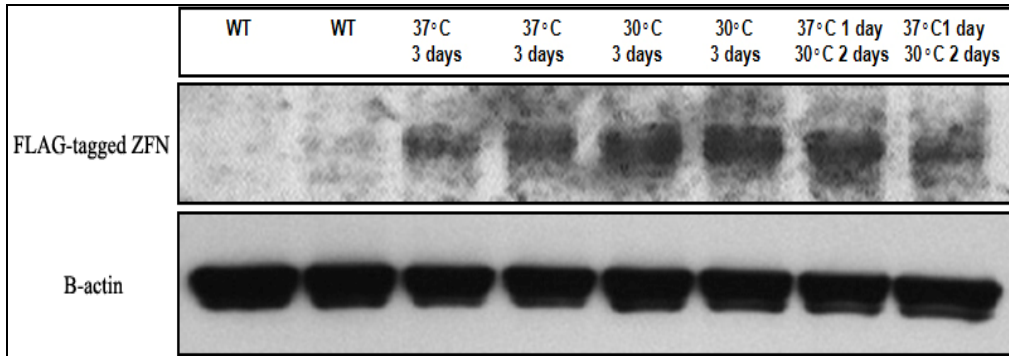
Figure 32 **Immunoblot detection of FLAG-tagged ZFN protein in transfected CLECs.** CLECs transfected with a plasmid vector encoding both left and right ZFNs were incubated at either 37°C or in mild hypothermia (30°C) for the indicated number of days. Cell lysates were analyzed (in duplicate) for levels of (**top**) FLAG-tagged ZFN proteins by immunoblotting with a monoclonal anti-FLAG antibody, rabbit anti-mouse IgG–horse radish peroxidase conjugate, detected with a chemiluminescence substrate and imaged with X-ray film on a Kodak processor. Untransfected CLECs (WT) were negative for expression of FLAG-tagged ZFN protein. (**Bottom**) Immunoblot for β-actin served as loading control.

2.3.2.3.3.   Site-specific endonuclease activity at AAVS1 locus in CLECs

Having demonstrated the expression of AAVS1 ZFNs (mRNA and protein) in CLECs following transient transfection with ZFN-encoding plasmid DNA, we next sought evidence for site-specific endonuclease activity of the expressed AAVS1 ZFNs. CLECs were transiently transfected with either two separate plasmid constructs (each encoding left or right AAVS1 ZFN; hereafter termed single constructs) or a single dual plasmid DNA encoding both left and right ZFNs (hereafter termed dual construct). Transfected CLECs were incubated at 37°C for 3 days or 37°C for 1 day followed by 30°C for 2 days, before genomic DNA was extracted for evaluation of AAVS1 site-specific genomic cleavage.

Genomic cleavage at AAVS1 locus and subsequent repair by the NHEJ pathway would be expected to form indels at sites of repair. We employed a *Cel*-1 mismatch nuclease assay (as detailed in Methods section 4.4.2.8.1) to screen PCR amplicons spanning the AAVS1 site for evidence of indels. *Cel*-1 nuclease only cleaves heteroduplex DNA. The presence of cleaved DNA products in the *Cel*-1 digested PCR amplicons (**Figure 33)** was evidence for indels at the AAVS1 locus of CLECs electroporated with AAVS1 ZFNs. The data indicated that the expressed ZFNs were fairly active in CLECs and induced site-specific genomic cleavage at the AAVS1 locus and repair by the NHEJ pathway. Based on densitometric measurements of cleaved and uncleaved PCR products, an estimate of the percentage of mutant cells (a measure of genomic cleavage efficiency) for CLECs treated with two single construct ZFNs at 37°C was 29.08 ± 3.70%; two single construct ZFNs at 30°C was 35.03 ± 1.83%; dual construct ZFN at 37°C was 37.16 ± 4.12% and dual construct ZFN at 30°C was 43.01 ± 1.93% (Data are mean ± SEM, n=3). A significantly greater proportion (P = 0.045) of mutant cells was observed only for CLECs treated with dual construct ZFN compared to two single construct ZFNs and incubated at 30°C. Differences in the proportion of mutant cells obtained with all other treatment combinations and conditions were statistically insignificant (P>0.05).
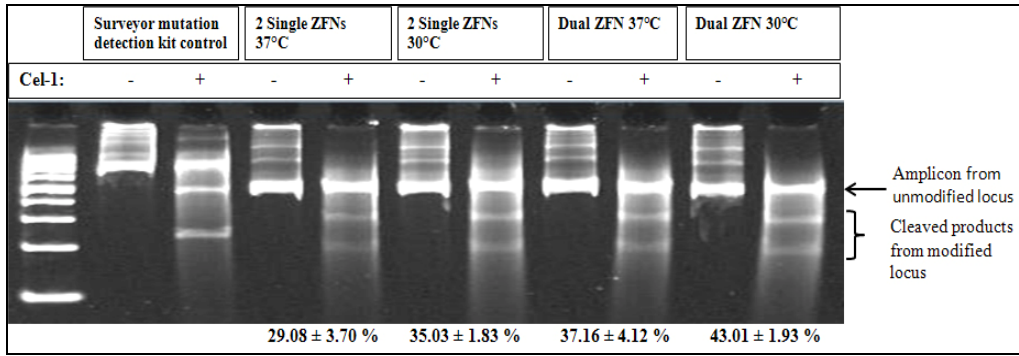
Figure 33   Site-specific cleavage and repair of AAVS1 genomic locus in CLECs. A genomic region spanning the AAVS1 ZFN target site was amplified from CLECs transiently electroporated with either two separate plasmid constructs encoding left or right AAVS1 ZFNs (2 single ZFNs) or a single dual construct plasmid encoding both left and right AAVS1 ZFNs (Dual ZFN). Cells were incubated at either 37°C for 3 days (37°C) or 37°C for 1 day followed by 30°C for 2 days (30°C). *Cel*-1 nuclease digested (+) or undigested (-) PCR amplicons were electrophoretically resolved in a 10% polyacrylamide gel, imaged and quantified using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. A PCR amplicon provided by the SURVEYOR® mutation detection kit served as a positive control for *Cel*-I nuclease digest. Estimates of the proportion of mutant cells (ZFN cleaved and repaired with indels) in the bulk treated population based on densitometric measurements are reported below respective gel images for each of the samples.

2.3.2.4. Optimization of site-specific integration in CLECs

One of the key goals of the project was to study the potential for utilizing AAVS1 ZFNs to mediate site-specific integration of a transgene of interest into the genome of CLECs. A major effort was therefore invested in optimizing conditions for site-specific integration of donor DNA at the AAVS1 locus in CLECs. Initial optimization experiments attempted to integrate a 50-bp donor DNA at the AAVS1 locus. Subsequently, we investigated site-specific genomic integration of two transgene cassettes, 4-kb (EGFP driven from CMV promoter) and 9-kb (hybrid FVIII cDNA driven from hFer promote) in size.

2.3.2.4.1.   Integration of 50-bp DNA in CLECs

CLECs co-electroporated with pZDonor (consisting of 50-bp donor DNA flanked by 800-bp sequences homologous to the AAVS1 genomic locus) and AAVS1 ZFN enhanced Sharkey variant, using conditions optimized for nucleofection with a Nucleofector™ I device (program T23) (**Figure 34**), were analyzed for evidence of site-specific integration by RFLP assay and junction PCR analysis.

Junction PCR showed positive amplification with a pair of vector specific and genomic specific primers only in CLECs co-electroporated with donor and ZFN plasmids (Donor + AASV1 ZFN) but not with donor plasmid alone (Donor only) (done in triplicate). Control genomic PCR served to show the integrity of genomic DNA and similar PCR amplification efficiency for both samples analyzed.

Similarly, RFLP assay (done in duplicate) was also positive for the diagnostic pattern of the cleaved amplicon, indicative of site-specific integration of the 50-bp donor DNA, only in CLECs co-treated with ZFNs (Donor + AAVS1 ZFN) but not when treated with donor only. Based on densitometric measurements of RFLP assay products, an estimated 3.39 ± 0.1 % of treated CLECs (Data are mean $\pm$ SEM, n=2) achieved site-specific integration of the 50-bp donor DNA.
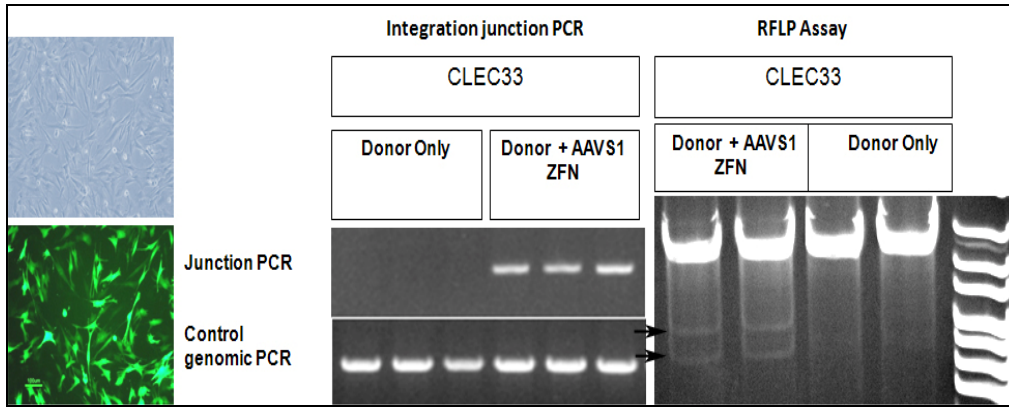
Figure 34 **ZFN-mediated site-specific integration into the AAVS1 locus of CLECs.** CLEC#33 electroporated with pEGFP-C1 and pZDonor with ("Donor + AAVS1 ZFN") or without ("Donor only") enhanced Sharkey plasmid was evaluated for gene transfer by fluorescence microscopy (**left**: original magnification x100) and for site-specific integration of a 50-bp donor DNA by integration junction PCR and RFLP assay. (**Center**) Integration junction PCR was performed (in triplicate) with a pair of vector specific and genome specific primers to amplify a 1 kb region spanning the integration junction. Control genomic PCR amplified a 900-bp region of the AAVS1 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. (**Right**) RFLP assay was performed (in duplicate) by digesting genome amplified PCR products (amplified with genome specific primers spanning the integration site) with *Hin*d III restriction enzyme followed by electrophoresis in a 5% polyacrylamide gel. Site-specific integration of 50-bp DNA results in insertion of a novel *Hin*d III restriction enzyme site which is evident by the presence of the two cleaved, faster migrating DNA bands (indicated by arrows). Densitometric measurements (performed with Quantity One software) of digested bands (modified AAVS1 locus) expressed as a percentage of total band intensities (modified + unmodified AAVS1 locus) gave the estimated site-specific integration frequency (reported in the text).

2.3.2.4.2.   Screening of different CLEC samples

    Having positively demonstrated the potential for ZFN-induced site-specific integrations in CLECs from a single donor sample (CLEC#33), we next tested CLEC samples from 6 different donors.

    Transfection efficiencies based on flow cytometric analysis of GFP-positive cells ranged from 22.4 to 71.3% across different samples (median transfection efficiency of 45.3%). Of a total of 7 CLEC samples tested, 4 were demonstrated by RFLP assay to be positive for site-specific integration of 50-bp donor DNA, following co-treatment with AAVS1 ZFNs (**Figure 35**). The percentage of donor DNA modified cells ranged from 12.1 to 22.2 % (median percentage of 16.5%, based on triplicate measurements for each sample). Transfection efficiency did not correlate with the percentage of genome-modified cells ($R^2 = 0.1$).

    The effect of mild hypothermia on the percentage of genome-modified cells was evaluated by RFLP assay for two CLEC samples (CLEC#33 and CLEC#16). Although a modest 1.2-fold increase ($P>0.05$) in the percentage of modified cells was observed for CLEC#33 when incubated at 30°C (13.0 ± 0.5%) compared to 37°C (10.6 ± 0.9%, Data are mean ± SEM, n=3), the same trend was not observed in another sample tested, CLEC#16. Based on our earlier results with K562 cells and greater ZFN protein levels detected in CLECs after mild hypothermia, we used mild hypothermic incubation for all subsequent ZFN treatments of CLECs.

    Although higher percentage of gene modification was observed with CLEC#35 (19.9 ± 0.5%) and #36 (22.2 ± 0.9%), all further experiments were done with CLEC#33 (13.0 ± 0.5%) which had greater proliferative capacity and offered ease of generating large numbers of cells for experiments.
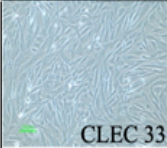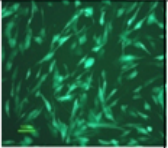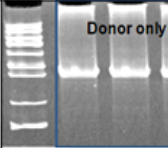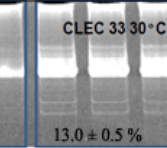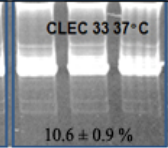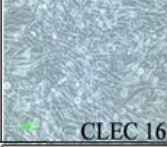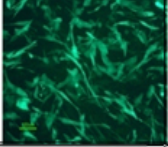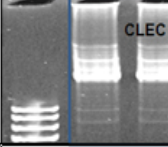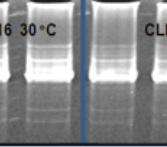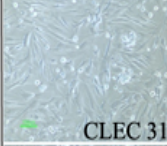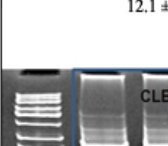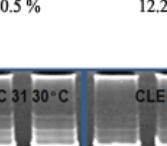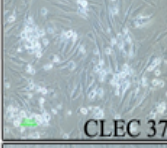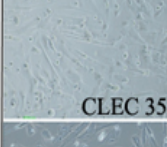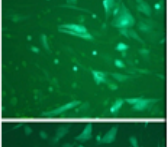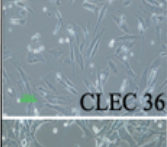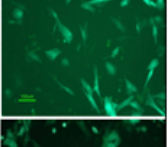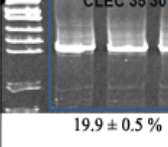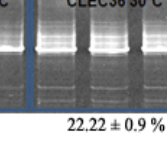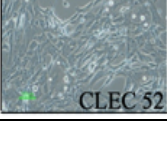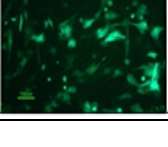
Figure 35 **Evaluation of different primary CLEC samples for ZFN-mediated site-specific integration of a 50-bp donor DNA into the AAVS1 locus.** CLECs isolated from different donor samples (#16, 31, 33, 35, 36, 37, 52) were co-electroporated with pEGFP-C1, pZDonor plasmid DNA with or without bicistronic AAVS1 ZFN variant (enhanced Sharkey) plasmid DNA. Transfected cells incubated at 37°C for 4 days or 37°C for 1 day followed by 30°C for 3 days (labeled as 30°C) were evaluated for GFP expression (1 day post-electroporation) by fluorescent microscopy and FACS analysis and for targeted integration of 50-bp donor DNA (4 days post-electroporation) by RFLP assay. (**Left panel**) Brightfield and fluorescence images of the indicated transfected CLEC samples (original magnification x100) and (**center panel**) percent GFP-positive cells as determined by FACS are shown (data are mean ± SEM; *n* = 3 per CLEC sample). **Right panel** shows RFLP assay products electrophoresed on 5% polyacrylamide gels, stained with ethidium bromide, imaged and quantified using BioRad®Gel Doc 2000 transilluminator and Quantity One software. Site-specific integration of the 50-bp donor DNA resulted in insertion of a *Hin*d III restriction enzyme site, cleavage at which generated two cleaved, faster migrating DNA bands. Densitometric measurements of digested bands (modified AAVS1 locus) expressed as a percentage of total band intensities (unmodified + modified AAVSI locus) gave the estimated site-specific integration frequency (shown below the gel images for each sample) or indicated as "Negative" where no site-specific integration was detected.

208

2.3.2.4.3.   Effect of donor DNA dose

In order to test if the percentage of modified cells could be increased by increasing donor DNA dose, we co-electroporated CLECs with increasing amounts of pZDonor DNA while fixing AAVS1 ZFN DNA dose at 5 μg and a reporter plasmid (EGFP) at 2 μg. We screened for site-specific integration by junction PCR and for overt signs of genotoxicity by monitoring if the percentage of GFP-positive cells drastically decreased after treatment (Day 4 *versus* Day 1). Flow cytometric analysis for phosphorylated histone H2AX served as an indicator of double-stranded DNA breaks[7], an indirect marker of genotoxicity (**Figure 36**).

Due to limited amounts of genomic DNA from treated cells and poor RFLP data, junction PCR products were normalized for PCR efficiency and expressed as a percentage of their respective control PCR amplicons. These data were used to compare site-specific integration efficiencies among groups. No junction PCR products were observed for samples treated with donor DNA only. Significantly greater normalized junction PCR amplification was observed in samples treated with 30 μg of donor DNA and ZFN (11.8 ± 0.3 %) compared with 20 μg (9.8 ± 0.2 %, $P = 0.0085$) or 10 μg (9.3 ± 0.3 %, $P = 0.01$) of donor DNA and ZFN.

The proportion of GFP-positive cells on day 4 expressed as a percentage of GFP-positive cells on day 1 post-treatment was used as an indicator of cellular toxicity. When compared with cells electroporated with pZDonor plasmid only (87.99 ± 0.42 %), significant decrease in GFP-positive cells was observed only in cells co-electroporated with ZFN plasmid and either 20 μg (80.79 ± 0.24 %, $P = 0.0001$) or 30 μg (72.2 ± 0.23 %, $P<0.0001$) of pZDonor plasmid DNA and not when co-electroporated with 10 μg of pZDonor plasmid (87.72 ± 0.92, $P = 0.80$, data are mean $\pm$ SEM, n=3). These results suggested that co-electroporation of ZFN plasmid resulted in significant cellular toxicity only when pZDonor plasmid DNA exceeded 10 μg.

Phosphohistone H2AX FACS analysis revealed that compared to donor only electroporated cells ($3.85 \pm .31$ %) there was a significant increase ($P<0.05$) in the percentage of cells with double-stranded DNA breaks in samples co-electroporated with ZFN irrespective of donor DNA dose. For co-electroporation groups, while no significant difference was observed between 10 μg (6.24 ± 0.37%) and 20 μg (6.85 ± 0.24 %) donor DNA dose, electroporation with 30 μg (8.87 ± 0.24 %) of donor DNA resulted in significant increase in phosphohistone H2AX-positive cells compared with both 10 μg and 20 μg donor DNA doses ($P<0.05$ for comparison of 30 μg donor DNA dose with 20 μg and 10 μg donor DNA dose). These results suggested that significant

DNA breaks could be induced when ZFN activity was combined with a high dose of donor DNA, i.e. 30 μg.

Taken together, although a slight but significant increase in gene targeting was observed with 30 μg donor DNA, significant toxicity was also demonstrated at this DNA dose. To minimize cellular toxicity, we performed co-electroporation of ZFN with amounts of donor DNA between 10 to 20 μg.
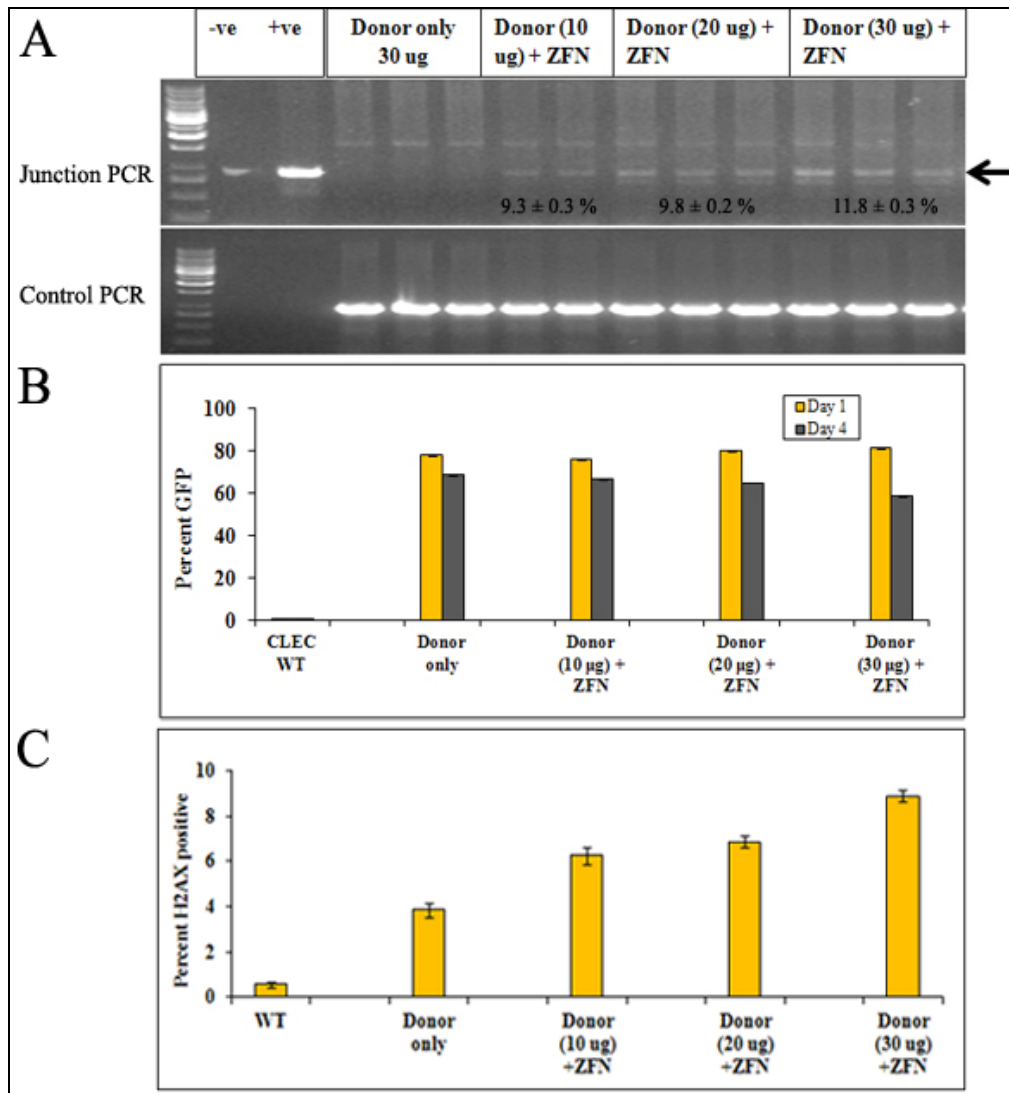
Figure 36   Investigation of donor DNA dose on gene targeting and cellular toxicity. Untreated CLECs (CLEC WT) and CLECs co-electroporated with 2 μg of pEGFP-C1 and varying doses of pZDonor plasmid DNA with (Donor + ZFN) or without (Donor only, 30 μg) a bicistronic AAVS1 ZFN variant (enhanced Sharkey) plasmid DNA (6 μg) were evaluated for gene transfer efficiency, targeted gene integration and for evidence of double-stranded DNA breaks. (**A**) Genomic DNAs extracted from CLECs 4 days post-electroporation were evaluated for site-specific integration of 50-bp donor DNA by integration junction PCR (all done in triplicate, except Donor (10 μg) + ZFN which was done in duplicate). "-ve" refers to minus template amplification while "+ve" refers to amplification of a K562 genomic DNA sample known to be positive for pZDonor site-specific integration. Control genomic PCR amplified a 900-bp region of the AAVS1 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Black arrow indicates the expected integration junction PCR amplicon. Densitometric measurements of junction PCR bands are given as percentage of the respective control PCR bands (indicated on gel image) [data are mean ± SEM, n=3 for all groups except "Donor (10 μg) + ZFN", n=2]. (**B**) Flow cytometric analysis for GFP-positive cells was done on days 1 and 4 post-electroporation. Compared to cells electroporated with pZDonor DNA only (Donor only), significantly fewer GFP-positive cells were observed on day 4 compared to day 1 for groups treated with 20 and 30 μg of pZDonor DNA and ZFN

211

(*P*<0.0001). Data are mean ± SEM; n= 3 per group. (**C**) Proportion of cells positive for phosphorylated histone H2AX, a marker for double-stranded DNA breaks, was quantified by flow cytometry following incubation with Phospho-Histone H2AX (Ser139) (20E3) rabbit mAb (Alexa Fluor® 647 conjugate). Compared to cells electroporated with donor DNA only, a significantly greater proportion of cells (*P*< 0.05) were phosphoH2AX-positive when co-electroporated with ZFN. Data are mean ± SEM; n= 3 per group.

2.3.2.4.4.   Effect of ZFN dose

The effect of increasing ZFN dose on targeted gene integration and cellular toxicity was investigated (**Figure 37**). CLECs were co-electroporated with pZDonor DNA (10 μg) and 2 μg of GFP reporter only or together with increasing doses of ZFNs (5 μg, 10 μg, 20 μg). Junction PCR demonstrated targeted integration of 50-bp donor DNA fragment only in samples co-electroporated with ZFNs. Band intensities of junction PCR amplicons were similar, suggesting that increasing amounts of ZFNs did not enhance the frequency of site-specific integration events. Control amplicons in all samples had comparable band intensities.

Genotoxicity and cellular toxicity were assessed by 3 different assays. Firstly, the proportion of GFP-positive cells remaining on day 4 expressed as a percentage of initial levels on day 1 was used as a measure of overt cell toxicity. Compared with cells treated with pZDonor DNA only (105.1 ± 0.18 %), significantly fewer ($P<0.0001$) GFP-positive cells were observed on Day 4 after treatment with 5 μg (88.37 ± 1.14 %), 10 μg (78.89 ± 0.54 %) and 20 μg (63.82 ± 0.66 %) of ZFN. Increasing ZFN dose from 5 μg to 10 μg ($P = 0.0017$) and to 20 μg ($P<0.0001$) resulted in significant reduction in GFP-positive cells on day 4. Secondly, the percentage of cells positive for phosphorylated histone H2AX determined by flow cytometry was used as a measure of DNA damage. Compared to cells electroporated without any DNA (1.09 ± 0.12 %) or with pZDonor DNA only (1.38 ± 0.13 %), significant increases ($P<0.05$) in H2AX-positivity were detected in cells co-electroporated with 5 μg (9.34 ± 0.14 %), 10 μg (9.87 ± 0.32 %) and 20 μg (14.19 ± 0.17 %) of ZFNs. Electroporation of CLECs with 20 μg of ZFN resulted in significantly increased ($P<0.05$) phosphorylated H2AX-positive cells compared with the two lower ZFN doses. The third assay, an MTS assay based on colorimetric reduction of a tetrazolium compound by mitochondrial dehydrogenase activity was used as an indirect indicator of cell viability. An equal number of cells from each treatment group was seeded and absorbance at 490 nm was measured after incubation with the tetrazolium compound and an electron coupling reagent. A standard curve of absorbance readings against different cell numbers yielded a linear and positive correlation ($R^2 = 0.9776$). Thus, raw absorbance readings were used in comparisons between groups. Compared with cells electroporated with donor DNA only (0.632 ± 0.013), significant decreases ($P<0.0001$) in absorbance readings were noted for cells co-electroporated with ZFNs at all doses (5 μg: 0.433 ± 0.017; 10 μg: 0.446 ± 0.009; 20 μg: 0.472 ± 0.013). Higher ZFN doses did not significantly alter cell viability.

Taken together, our results set the ZFN dose between 5 to 10 μg, and not exceeding 20 μg, to avoid overt cellular toxicity.
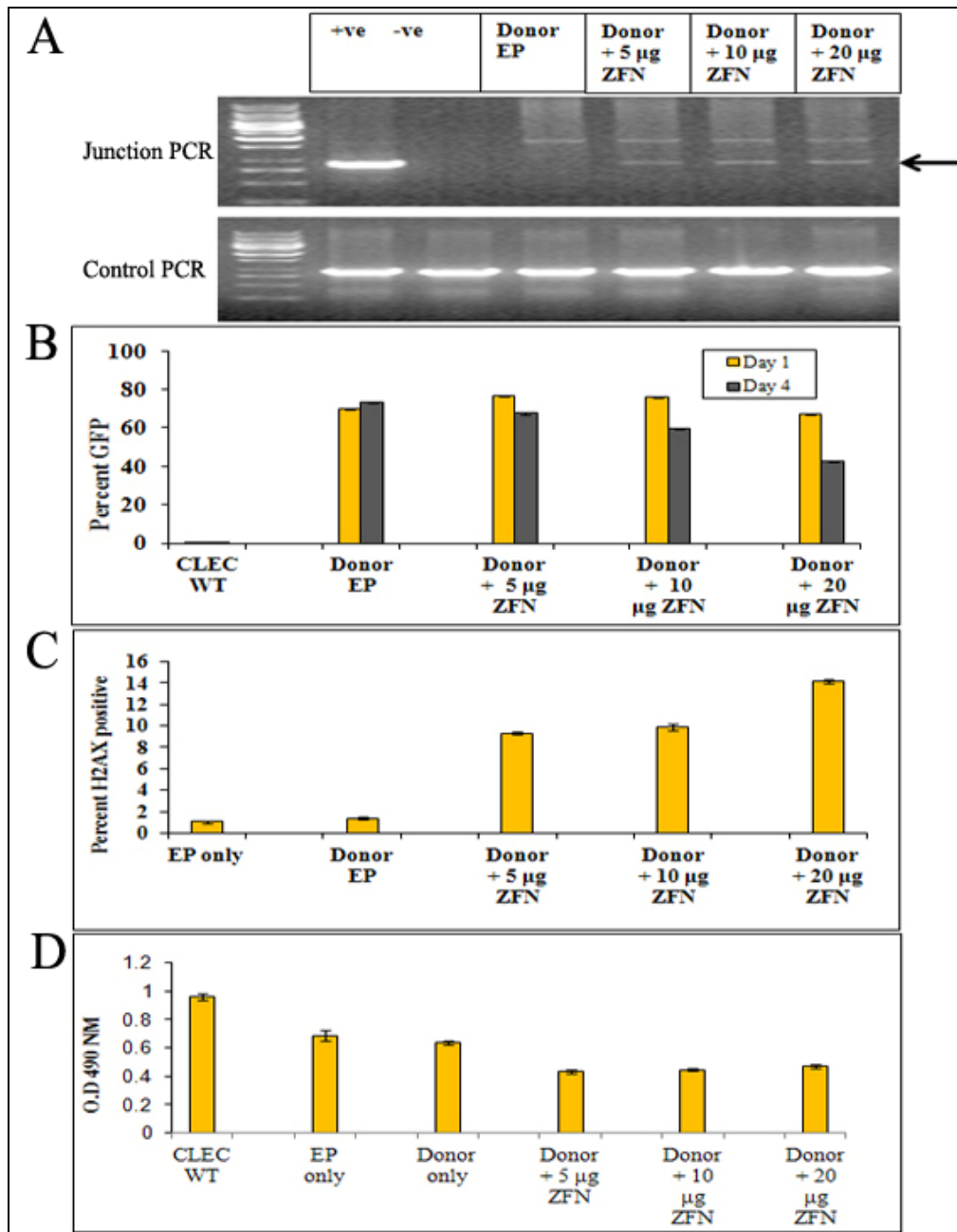
Figure 37   Effects of ZFN dose on gene targeting and cellular toxicity.
Untreated CLECs (CLEC WT) and CLECs co-electroporated with 2 μg of pEGFP-C1, 10 μg pZDonor plasmid (Donor EP) alone or with (Donor + ZFN) varying doses of a bicistronic AAVS1 ZFN variant (enhanced Sharkey) plasmid DNA (5, 10, 20 μg) were evaluated for gene transfer efficiency, targeted gene integration, cell viability and double-stranded DNA breaks. (**A**) Genomic DNA extracted from CLECs 4 days post-electroporation were evaluated for site-specific integration of 50-bp donor DNA by integration junction PCR. "-ve" refers to minus template amplification while "+ve" refers to amplification of a K562 genomic DNA sample previously identified to be positive for pZDonor site-specific integration. Control genomic PCR amplified a 900-bp region of the AAVS1 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Black arrow indicates the predicted integration junction PCR amplicons. (**B**) Flow cytometric analysis for GFP-positive cells on days 1 and 4 post-electroporation. Compared to cells electroporated with pZDonor DNA only (Donor

EP), significantly fewer GFP-positive cells were observed on day 4 compared to day 1 for groups treated with all doses of ZFN ($p<0.0001$). Data are mean ± SEM; n= 3 per group. (**C**) Proportion of cells positive for phosphorylated histone H2AX, a marker for double-stranded DNA breaks, was quantified by flow cytometry following incubation with Phospho-Histone H2AX (Ser139) (20E3) rabbit mAb (Alexa Fluor[®] 647 conjugate). Compared to cells electroporated with donor DNA only, a significantly greater proportion of cells ($p< 0.05$) were H2AX-positive when co-electroporated with ZFN. Data are mean ± SEM; n= 3 per group. (**D**) MTS assay (CellTiter 96[®]Aqueous One solution cell proliferation assay kit) (O.D. 490 nm) was performed on untreated CLECs (CLEC WT), CLECS electroporated with pZDonor only (Donor EP) or with pZDonor and varying doses of ZFNs, 1 day post- treatment. Data are mean ± SEM; n= 4 per group. Significant decrease in O.D. 490 nm was observed in cells electroporated with ZFN, compared to Donor only cells ($p<0.05$).

2.3.2.5. Integration of 4-kb EGFP cassette

Having successfully integrated a 50-bp donor DNA into the AAVS1 locus in CLECs following ZFN treatment, we next used optimized conditions to integrate a 4-kb donor DNA fragment consisting of an EGFP reporter gene driven from a CMV promoter and a neomycin resistance selection marker.

CLECS were co-electroporated with 15 µg of pZDonor EGFP and 7 µg of bicistronic plasmid encoding for both left and right AAVS1 enhanced Sharkey ZFNs using T20 and CM102 programs of the Amaxa® Nucleofector™ I and 4D devices, respectively. Stably transfected cells, selected by G418 resistance, were screened by left and right integration junction PCR on extracted genomic DNA (**Figure 38**). CLECs electroporated with CM102 program were positive for both left and right integration junction PCR while electroporation with T20 program resulted in only a positive right integration junction PCR. The absence of the left integration junction in these cells raised the possibility of partial integration of the transgene cassette, possibly due to incomplete homologous recombination-mediated repair. However, as our previous results had already shown a lower efficiency for amplifying the left integration junction, a negative left junction PCR amplicon could have resulted from a lower integration efficiency when T20 program was used. Sequencing of the integration junction PCR amplicons showed precise integration of the donor DNA (without indels) adjacent to the AAVS1 site cleaved by ZFNs.

The demonstration of left and right integration junctions by PCR was taken as evidence of precise integration a 4-kb donor DNA into the AAVS1 locus mediated by ZFN treatment.
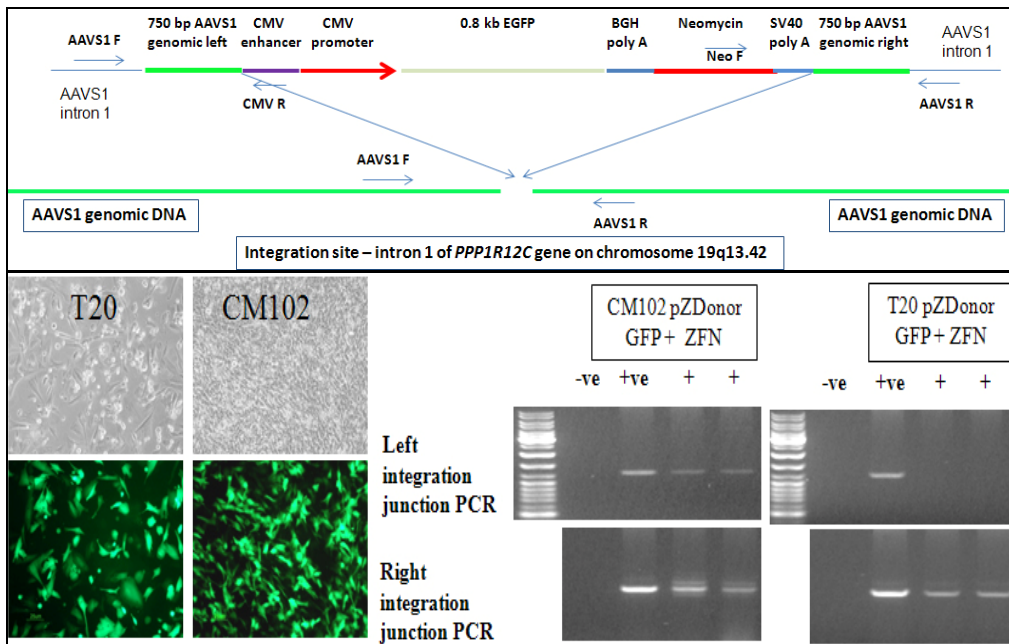
Figure 38　Integration of pZDonor EGFP into AAVS1 locus of CLECs.
(Top) Schematic showing homologous recombination-mediated integration of
pZDonor EGFP vector into AAVS1 locus and the primer sets (right junction: AAVS1
Forward; CMV Reverse, left junction: Neo Forward; AAVS1 Reverse) used for
integration junction PCR analysis of site-specific integration. (**Bottom**) CLEC#33
co-electroporated with pZDonor EGFP plasmid DNA (4-kb EGFP cassette) and with
a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs
(+) were selected with G418. Genomic DNA extracted from stably transfected cells
was analyzed by integration junction PCR for evidence of donor DNA integration at
the AAVS1 locus. CM102 and T20 refer to two different Amaxa electroporation
programs used to electroporate CLECs. "-ve" refers to minus template PCR
amplification, while "+ve" refers to amplification of a K562 genomic DNA sample
previously identified to be positive for AAVS1 site-specific integration of the 4-kb
GFP cassette. (**Right**): Left and right integration junctions were amplified with vector
specific and genome specific primers while control genomic PCR was performed
with a pair of primers amplifying a 900-bp region of the AAVS1 locus (all done in
duplicate). Amplified products were electrophoresed on 1% agarose gels and imaged
using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. (**Left**):
Representative brightfield and fluorescence images taken 1 day post-electroporation
(Original magnification x100).

2.3.2.6.   Integration of 9-kb hybrid FVIII cassette

Site-specific integration of a hybrid FVIII cassette together with a neomycin resistance gene was one of the main goals of our project. Having demonstrated integration of 50-bp and 4-kb DNA donors, we were encouraged to attempt integration of a 9-kb hybrid FVIII transgene into the AAVS1 locus of CLECs (**Figure 39**).

CLECs electroporated with pZDonor hybrid FVIII only or together with AAVS1 enhanced Sharkey ZFNs using programs T23, T20 (with a Nucleofector™ I device) or CM102 (with Nucleofector™ 4D device) were selected with G418. Stably transfected CLECs derived from electroporating pZDonor hybrid FVIII plasmid only without any ZFNs (CM102 FVIII only) were negative for site-specific integration. G418-resistant cells derived from electroporation with program T20 were positive for both integration junctions by junction PCR analysis, while cells electroporated with CM102 were only positive for the right integration junction. Positive amplification of the left junction from T20 but not from CM102 treated CLECs raised two possibilities. The first was that the efficiency of site-specific integration attained with CM102 was not sufficient to allow PCR detection of the left integration junction which we knew to be less efficiently amplified compared to the right integration junction. The other possibility was that under certain circumstances, incomplete or only partial integration of the transgene cassette was achieved, resulting in only the formation of the right integration junction but not the left. Stable CLECs derived from electroporation with T23 program failed to show any evidence of site-specific integration (negative for both left and right junction PCR) suggesting that for that particular experiment, homologous recombination-mediated transgene integration was not achieved at all. Amplification from a control genomic locus served as the positive control for PCR and was comparable among all groups.

Collectively, these results suggest that although electroporation conditions and transfection efficiencies are important factors for successful ZFN induction of site-specific genomic cleavage, subsequent homologous recombination-mediated transgene integration also requires an intrinsic ability of cells to mediate DNA repair; and this latter capacity could possibly be a limiting factor. The ability of cells to repair DNA breaks and integrate transgenes by homologous recombination could also be influenced by other factors such as the cell cycle status at the point when DNA damage should occur. Such factors may be crucial determinants of successful integration of the whole transgene cassette in ZFN-treated CLECs.
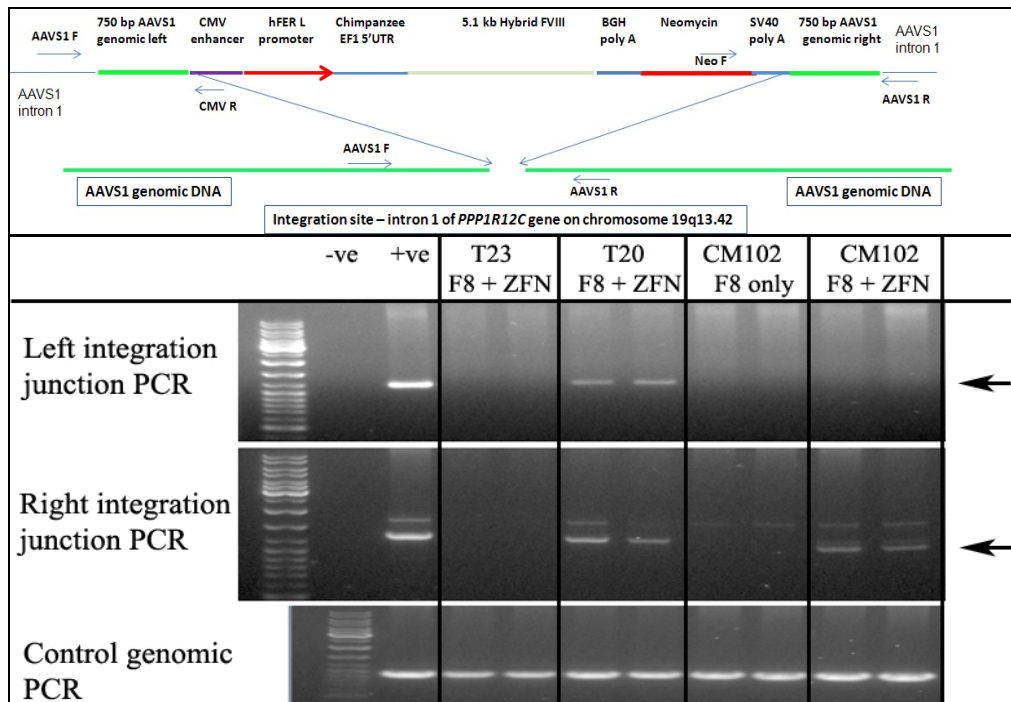
Figure 39   Integration of pZDonor hybrid FVIII into AAVS1 locus of CLECs. (Top) Schematic showing homologous recombination-mediated integration of pZDonor hybrid FVIII vector into AAVS1 locus and the primer sets (right junction: AAVS1 Forward; CMV Reverse; left junction: Neo Forward; AAVS1 Reverse) used for integration junction PCR analysis of site-specific integration. **(Bottom)** CLEC#33 co-electroporated with pZDonor hybrid FVIII donor plasmid DNA (9-kb hybrid FVIII cassette) with (+ ZFN) or without a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were selected with G418. Genomic DNA extracted from the resulting stable cell populations were analyzed by integration junction PCR for evidence of donor DNA integration at the AAVS1 locus. T23, T20 and CM102 refer to 3 different Amaxa electroporation programs used. "-ve" Refers to minus template amplification while "+ve" refers to amplification of a K562 genomic DNA sample previously identified to be positive for AAVS1 site-specific integration of the 9-kb hybrid FVIII cassette. Left and right integration junctions were amplified with vector specific and genome specific primers while control genomic PCR was performed with a pair of primers amplifying a 900-bp region of the AAVS1 locus (all done in duplicate). Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Black arrows indicate the predicted integration junction PCR amplicons.

220

2.3.2.7. Durable FVIII secretion from site-specific integration of FVIII transgene at the AAVS1 locus

Having demonstrated the potential for site-specific integration of the FVIII transgene into the AAVS1 locus, we proceeded to monitor the durability of FVIII secretion from a bulk population of CLECs that had evidence of site-specific FVIII transgene integration.

CLECs co-electroporated with pZDonor hybrid FVIII donor plasmid DNA (9-kb hybrid FVIII cassette) and a bicistronic plasmid encoding for left and right AAVS1 enhanced Sharkey ZFNs, with or without G418 selection, were examined for evidence of AAVS1-site specific integration 3 weeks post-electroporation. We demonstrated site-specific integration of hybrid FVIII transgene in CLECs samples that had been selected with G418 but not in unselected CLECs. Left and right integration junction PCR (**Figure 40A**) showed correct amplicons consistent with integration of the hybrid FVIII donor into the AAVS1 locus in G418-selected CLECs. In contrast, the inability to amplify both integration junctions in unselected CLECs suggested that very few or no cells had sustained site-specific integrations. Amplification of a control locus from both G418 selected and unselected cells were clearly and similarly positive in both groups, showing that genomic DNA samples were of comparable quality.

Consistent with stable and site-specific integration of FVIII transgene in G418-selected CLECs, comparable FVIII expression (day 6: 357.18 ± 1.85 mUnits/ 10^6 cells/24 hr; day 23: 278.09 ± 7.55 mUnits/ 10^6 cells/24 hr; day 30: 317 ± 9.23 mUnits/ 10^6 cells/24 hr) was evident from FVIII activity in conditioned media of treated cells from days 6 to 30 post-electroporation (**Figure 40B**). By way of contrast, CLECs that were not selected with G418 and in which electroporated plasmid was nearly all episomal had greatly reduced FVIII levels (day 6: 357.18 ± 1.85 mUnits/ 10^6 cells/24 hr; day 23: 15.45 ± 0.875 mUnits/ 10^6 cells/24 hr) by day 23 post-electroporation.

Collectively, these data showed that ZFN-mediated FVIII transgene integration into the AAVS1 locus conferred durable FVIII expression in CLECs. However, to exclude the possibility that random integration of FVIII transgene was the source of sustained FVIII expression, we undertook further analysis on a clonal population of CLECs to confirm that durable FVIII expression originated from AAVS1-specific integration of the FVIII transgene.
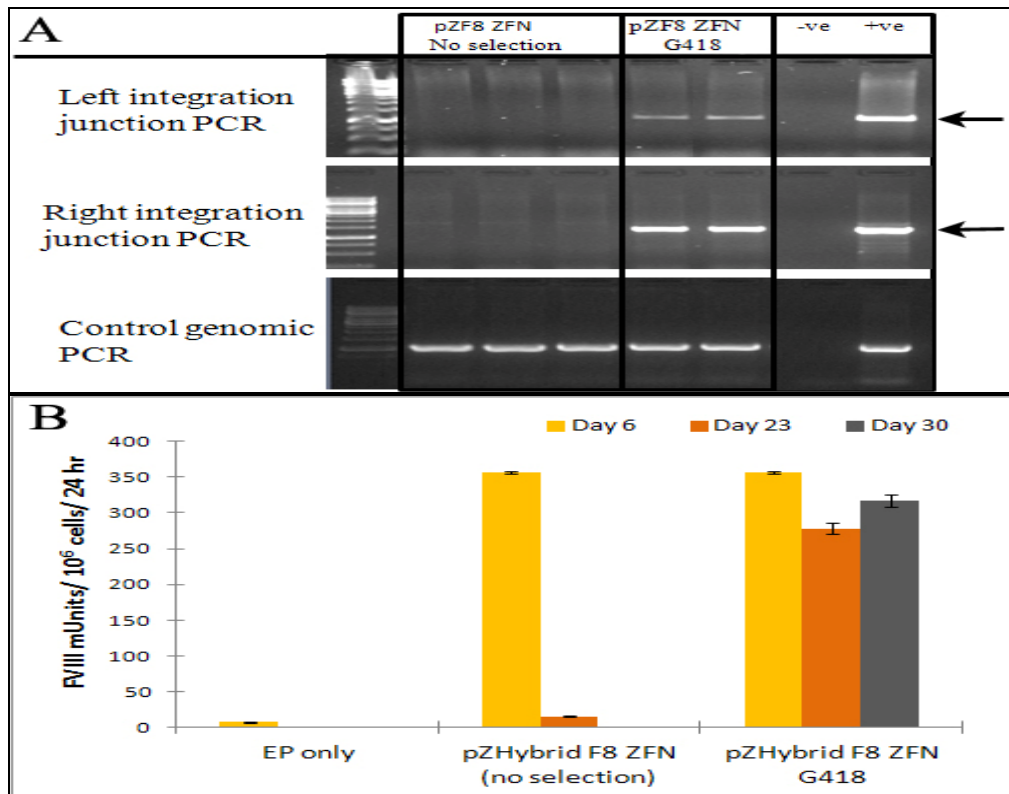
Figure 40 Evidence of ZFN-mediated site-specific integration of 9-kb donor DNA into the AAVS1 locus of CLECs correlates with durable FVIII secretion. (**A**) CLEC#33 co-electroporated using program CM113 with pZDonor hybrid FVIII donor plasmid (9-kb hybrid FVIII cassette) and a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were either unselected (no selection) or selected with G418. Genomic DNA extracted from the resulting cell populations was analyzed by integration junction PCR for evidence of donor DNA integration at the AAVS1 locus. "-ve" Refers to minus template amplification, while "+ve" refers to amplification of a K562 genomic DNA sample previously identified to be positive for AAVS1 site-specific integration of the 9-kb hybrid FVIII cassette. Left and right integration junctions were amplified with vector specific and genome specific primers. Control genomic PCR was performed with primers amplifying a 900-bp region of the AAVS1 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Black arrows indicate the expected integration junction PCR amplicons. (**B**) CLECs electroporated without any plasmid (EP only) or co-electroporated with pZDonor hybrid FVIII donor plasmid DNA and bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs with (G418) or without (no selection) a 7-day G418 selection beginning 7 days post-electroporation were analyzed for durable FVIII expression. FVIII activity in conditioned media from the indicated treatment groups assayed on days 6, 23 and 30 post- electroporation (Coamatic® FVIII assay, Chromogenix) are shown. Data are mean ± SEM; n= 3 per group, except for day 6 samples from pZHybrid F8 ZFN no selection group, where n=1.

2.3.2.8.   Oligoclonal cells with partial integration of 9-kb hybrid FVIII cassette

A puzzling phenomenon of our study of ZFN-mediated modification of CLECs with pZDonor hybrid FVIII vector were occasions when only a single integration junction, often the right junction, could be detected in the bulk CLEC population (**Figure 41**). To delve into these intriguing observations, we screened oligoclonal populations for the presence of both left and right integration junctions to determine if there had been incomplete integration of the 9-kb donor DNA.

We performed direct PCR to detect the right integration junction at the AAVS1 locus on 72 G418-resistant oligoclonal CLECs, of which 22 were positive (**Figure 42**). Further screening of positively amplified right junction PCR amplicons by a nested PCR confirmed that of the 22 positive clones, only 13 were positively by nested PCR. Surprisingly, all 72 clones were negative for the left integration junction.

We investigated if all the essential components of the integrated donor (Hfer promoter, FVIII cDNA, poly A sequences and neomycin resistance gene) necessary for FVIII expression were present in 5 selected clones (clone#1, 10, 36, 40 and 48) that were right junction positive but left junction negative. Direct PCR was used to screen *in situ* lysed clonal cells for presence of (1) AAVS1 Left junction and hfer promoter; (2) FVIII porcine A1 domain; (3) FVIII human A2 and B-domains; (4) FVIII porcine A3 domain; (5) FVIII human C1 and C2 domains; and (6) neomycin resistance gene cDNA, as depicted and numbered in **Figure 43A**. The results show that while all the above donor DNA components were present in CLECs positive for both integration junctions, all 5 selected clones had very few donor components, except for neomycin cDNA (labeled as -6) and perhaps FVIII C1 and C2 domain (labeled as -5) (**Figure 43B**), confirming that the presence of a single junction by PCR was indicative of partial integration of donor DNA. Such partial integration would be expected to induce no FVIII expression. Consistent with this, FVIII activity was very low or undetectable in such clonal CLEC populations. This also explained the drastic reduction in FVIII expression by day 42 post-electroporation in G418-resistant but unscreened bulk populations of pZDonor Hybrid FVIII and ZFN-treated cells (**Figure 41B**), unlike stably modified cells that were screen positive for both left and right integration junctions (**Figure 40B**).

While the actual reasons for incomplete or partial vector integration were unclear, positive results from earlier experiments (section 2.2.2.8) suggest that initial bulk screening of ZFN-treated cells for evidence of both left and right integration junctions is useful for identifying desirable oligoclones with complete transgene

integration in future. Factors that were difficult to control such as cell cycle status at time of DNA damage repair or even the manner in which Holliday junctions were resolved could have contributed to the incomplete transgene integration phenomenon that we encountered in some instances.
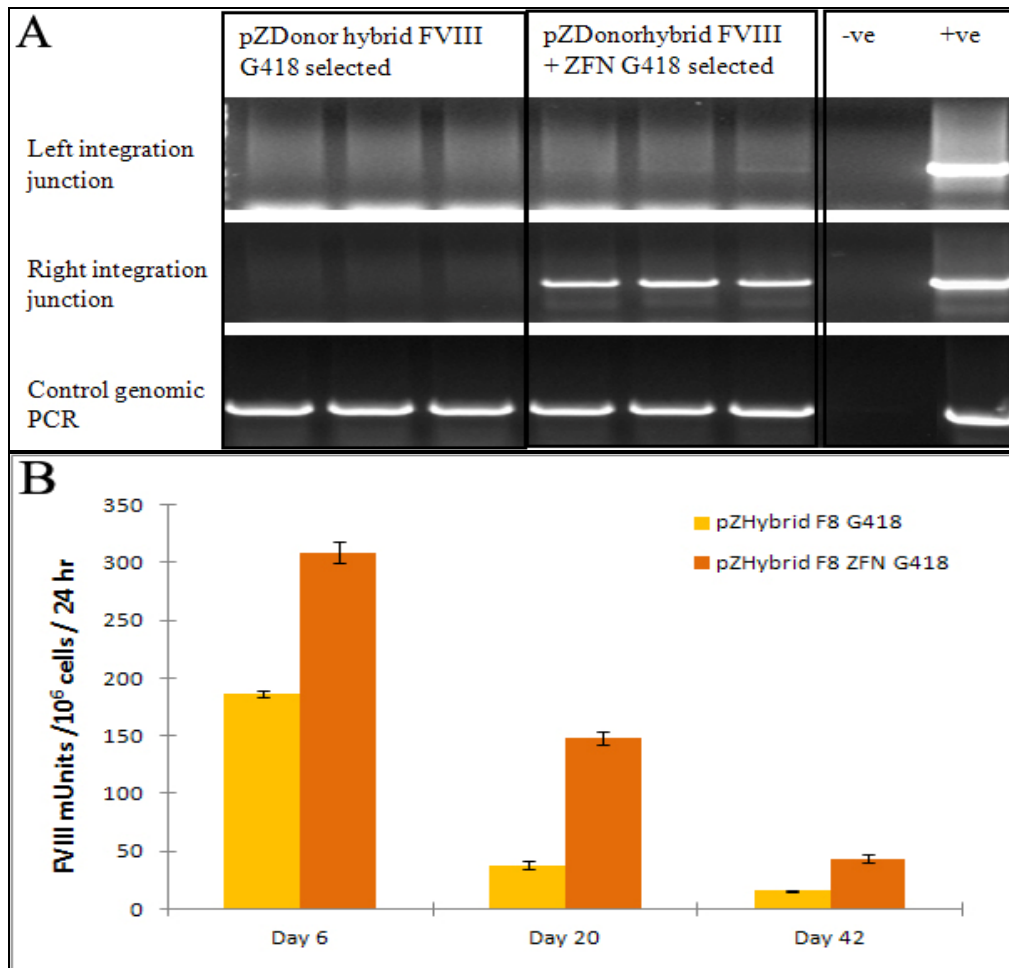
Figure 41　　　Bulk CLEC population which was positive only for the right integration junction following integration of pZDonor hybrid FVIII vector. (**A**) Genomic DNA extracted from a bulk population of G418-selected CLECs treated with either pZDonor hybrid FVIII only or in combination with AAVS1 ZFNs (pZDonor hybrid FVIII + ZFN) were analyzed by PCR for the presence of left and right integration junctions (done in triplicate). Control genomic PCR amplified a 1-kb region close to the AAVS1 locus on chromosome 19q13.42. Amplification products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator. Negative control (-ve) was amplification from a minus template PCR while positive control (+ve) was amplified from genomic DNA of K562 cells with a known complete integration of pZDonor hybrid FVIII. (**B**) FVIII activity in conditioned media of a bulk population of G418-selected CLECs electroporated with pZDonor hybrid FVIII with or without ZFNs was assayed on days 6, 20 and 42 post-electroporation. Data are mean ± SEM, n=3 per group.
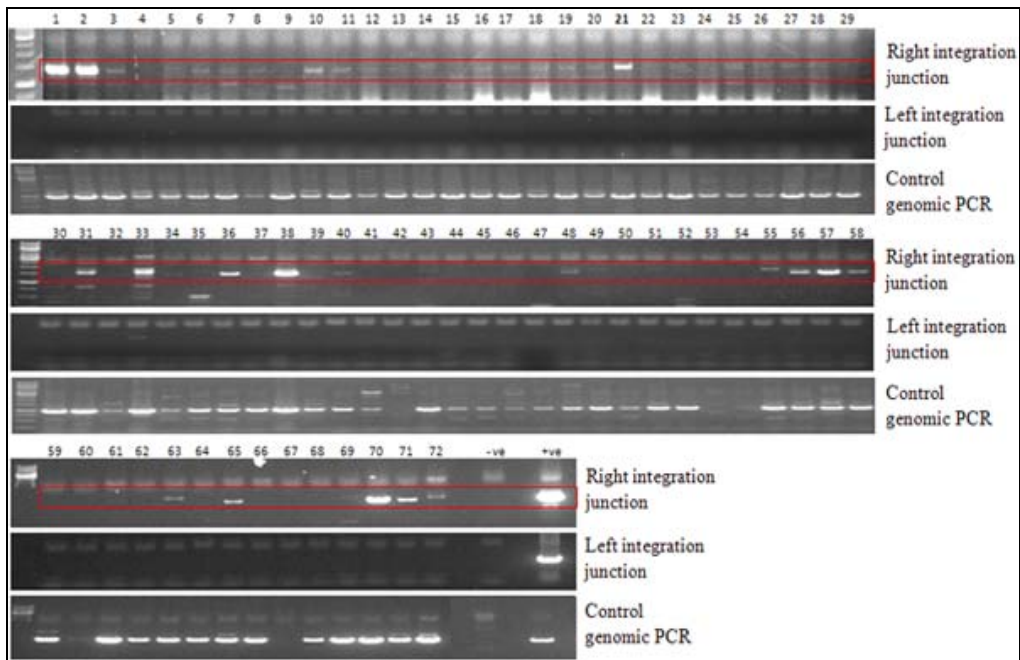
Figure 42 **Screening of oligoclonal CLECs for AAVS1 site-specific integration of pZDonor hybrid FVIII vector.** Flow sorted oligoclonal (4 cells per well) populations of G418-selected CLECs treated with pZDonor hybrid FVIII and ZFN were lysed *in situ* and analyzed for left and right integration junctions and a control genomic locus (1 kb region close to chromosome 19q13.42) using Phusion® Human Specimen Direct PCR kit. Amplification products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator. Negative control (-ve) was amplification from a minus template PCR while positive control (+ve) was amplification from genomic DNA of K562 known to have a complete integration of pZDonor hybrid FVIII. Red rectangles demarcate the predicted position of the right junction amplicon.
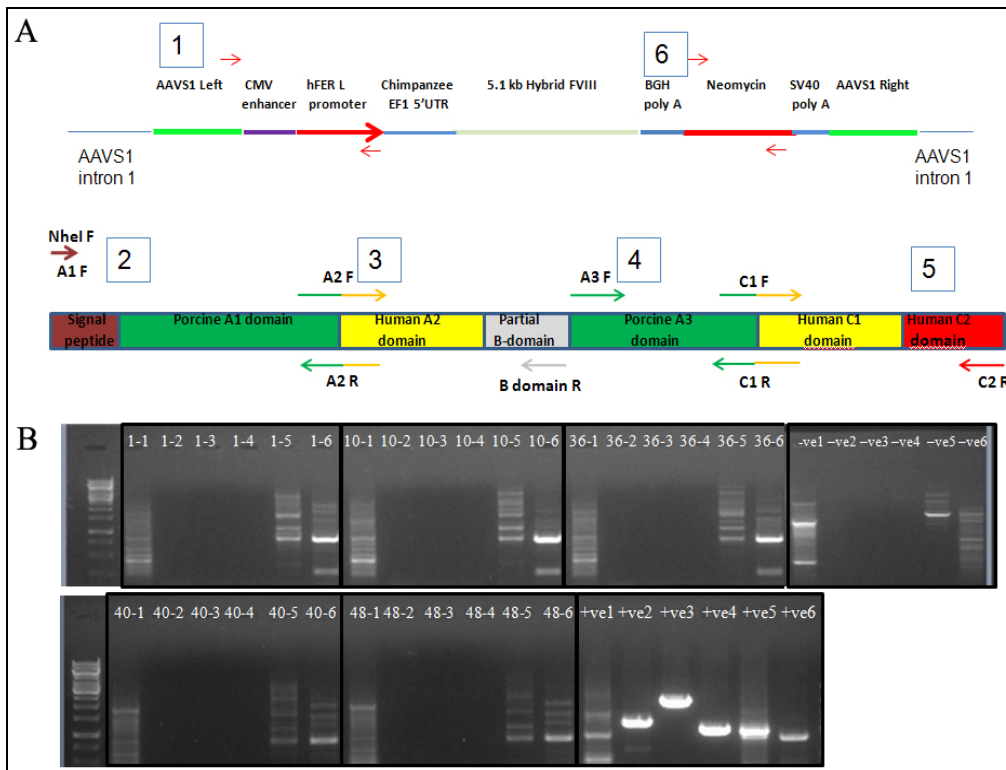
Figure 43   PCR analysis of selected oligoclonal CLECs for completeness of integration of pZDonor hybrid FVIII vector. (**A**) Schematic illustrating hybrid FVIII transgene integrated at the AAVS1 locus (**top**) and the domains of hybrid FVIII cDNA (**bottom**). Arrows and boxed numbers highlight regions that were investigated by PCR. PCR reaction #1 amplified from AAVS1 genomic region (5' ttcgggtcacctctcactcc3') to the end of human ferritin promoter (5'ttatggtgcgccggccctcg3'); reaction #2 amplified from the signal peptide (5'gccgctagcgatgcaaatagagctctcca3') to the end of FVIII porcine A1 domain (5'aggatgcttcttggcaactgagcggatttggataaaggggaga3'); reaction #3 amplified from the start of FVIII human A2 domain (5'tctccctttatccaaatccgctcagttgccaagaagcatcct3') to the end of partial B-domain (5'gcgggggctctgattttcatcctc3'); reaction #4 amplified FVIII porcine A3 domain (5'agctttcagaagagaacccgacac3'; 5'tcccaggggagtctgacacttcttgctgtacaccaggaaagt3'); reaction #5 amplified from FVIII human C1 domain (5'actttcctggtgtacagcaagaagtgtcagactcccctggga3') to the end of C2 domain (5'agtgctagctcagtagaggtcctgtgcc3'); and reaction #6 amplified neomycin resistance cDNA (5'ttgcacgcaggttctccggc3'; 5'ggcgtcgcttggtcggtcat3'). (**B**) Selected clones (#1, 10, 36, 40, 48) of G418-selected CLECs treated with pZDonor hybrid FVIII and ZFN were investigated by direct PCR for the presence of the different segments of the integrated transgene with primer pairs (shown in the schematic) as indicated by hyphenated numbers. Negative control (-ve) was amplification from a minus template PCR while positive control (+ve) was amplified from genomic DNA of K562 cells known to have completely integrated pZDonor hybrid FVIII. Amplification products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator.

2.3.3. A promoter trap strategy for site-specific transgene integration

In our initial exploration, cells with stable transgene integration were selected on the basis of G418 resistance conferred by expression of neomycin resistance gene driven from a constitutive exogenous promoter. However, such a selection strategy inevitably resulted in contamination by cells with random transgene integrations amongst a minor population of cells with the desired AAVS1 site-specific integration, thereby complicating downstream evaluations of the accuracy of site-specific transgene integration, biosafety and durability of transgene expression. We attempted to reduce the admixture of cells with random and site-specific integrations using a promoterless puromycin resistance gene plasmid construct with a splice acceptor site and self-cleaving 2A peptide sequence whereby puromycin resistance gene would be expressed only when driven off an endogenous promoter in the event of integration close to such a promoter. As the AAVS1 ZFN-induced genomic cleavage site is close to the endogenous *PPP1R12C* promoter, we anticipated that the majority of cells with AAVS1 site-specific integration would express puromycin resistance expressed from the endogenous *PPP1R12C* promoter and thereby substantially enrich the population for the desired site-specific integration using puromycin selection. Using this strategy, we tested for ZFN-mediated integration of pAAVS1-SA-2A-puro-pA donor[1] (1-kb donor insert), pAAVS-CAGGS-EGFP[1] (4.2-kb donor insert) and pAAVS1-SA-2A-puro-Hybrid FVIII (9-kb donor insert) vectors at the AAVS1 locus.

2.3.3.1. AAVS1 site-specific integration of a complete 1-kb puromycin cassette in CLECs

Stable cultures of puromycin resistant CLECs derived by co-electroporation with pAAVS1-SA-2A-puro-pA donor and a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were investigated for evidence of donor DNA integration at the AAVS1 locus. PCR amplification of both left and right integration junctions from genomic DNA isolated from puromycin-resistant stable CLECs provided strong evidence for site-specific integration of both the donor DNA vectors. Complete integration of pAAVS1-SA-2A-puro-pA donor (1 kb) at the AAVS1 locus was demonstrated by PCR amplification of the entire integrated vector (3 kb) using genome-specific primers overlapping the integration site (**Figure 44**). Comparatively, very weak amplification of the unmodified endogenous AAVS1 locus (Endogenous locus) suggested that the majority of puromycin-selected cells had biallelic integration of the puromycin cassette at the AAVS1 locus. Sequencing of the

amplified fragments verified complete donor integration and moreover showed no insertions or deletions at the integration junctions.
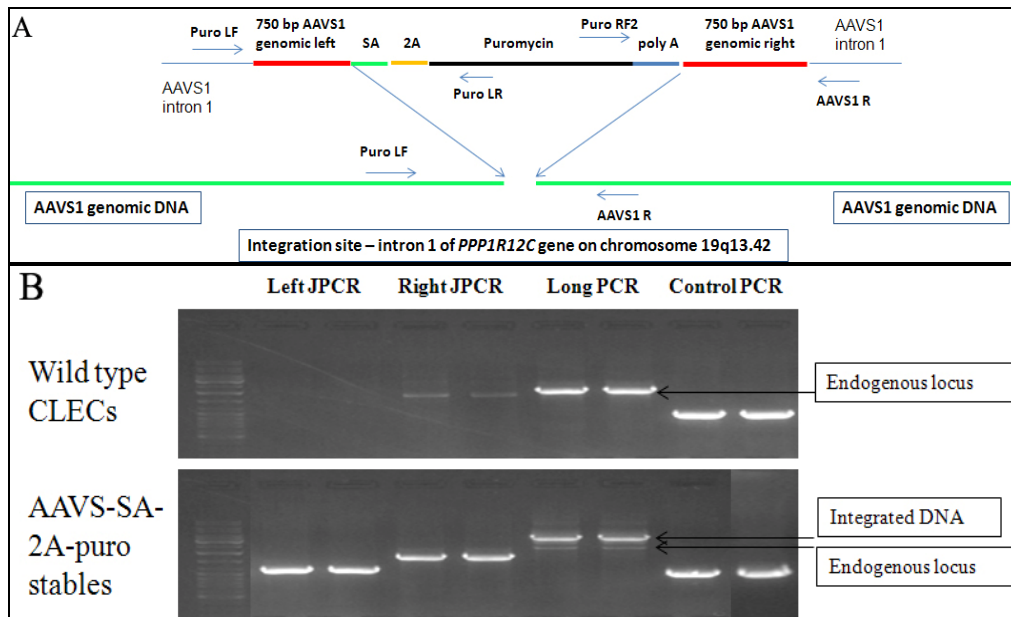
Figure 44   Complete integration of SA-2A-puro-pA donor into AAVS1 locus of CLECs. (A) Schematic showing homologous recombination-mediated integration of AAVS-SA-2A-puro-pA donor into AAVS1 locus and the primer sets (left junction: Puro LF; Puro LR, right junction: Puro RF2; AAVS1 R) used for integration junction and overlapping PCR (Puro LF; AAVS1 R) analysis of site-specific integration. (B) CLEC#33 co-electroporated with AAVS-SA-2A-puro-pA donor plasmid DNA (1 kb puromycin cassette) and with a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were selected with puromycin. Genomic DNA extracted from the resulting stable cell populations were analyzed by integration junction PCR and overlapping PCR for evidence of donor DNA integration at the AAVS1 locus. Left (Left JPCR; 1.1 kb) and right (Right JPCR; 1.6 kb) integration junctions were amplified with vector specific and genome specific primers as indicated above, while control PCR was performed with genome specific primers for a 900-bp region of the AAVS1 locus (all done in duplicate). PCR overlapping the integration site (Long PCR; amplicon of unmodified locus without integration/endogenous locus - 2 kb; amplicon of locus integrated with donor DNA-3kb) was performed with genome specific primers close to the integration site. Similar PCR performed on unmodified wild type CLECs served as controls for the unmodified locus. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Identity of junction PCR products and long PCR products were confirmed by sequencing.

2.3.3.2.   AAVS1 site-specific integration of complete 4-kb CAGGS EGFP cassette in CLECs

Integration of pAAVS-CAGGS-EGFP vector was shown by integration junction PCR as well as by two long PCR with vector specific and genome specific primer pairs (**Figure 45B**) covering most of the integrated transgene except a GC-rich 1-kb region within the transgene (CAGGS promoter) that could not be amplified. Amplified fragments were sequence verified and found not to have incurred deletions or insertions at integrations sites. CLECs with stable integration of EGFP cDNA at the AAVS1 locus expressed EGFP stably for up to 33 days in culture, when the experiment was terminated (**Figure 45C**). This showed that efficient selection and enrichment of cells with site-specific integration could be achieved using a promoter trap strategy with a promoterless construct. These experiments further demonstrate the capacity to integrate complete transgene inserts of up to 4.2 kb in size at the intended AAVS1 locus.
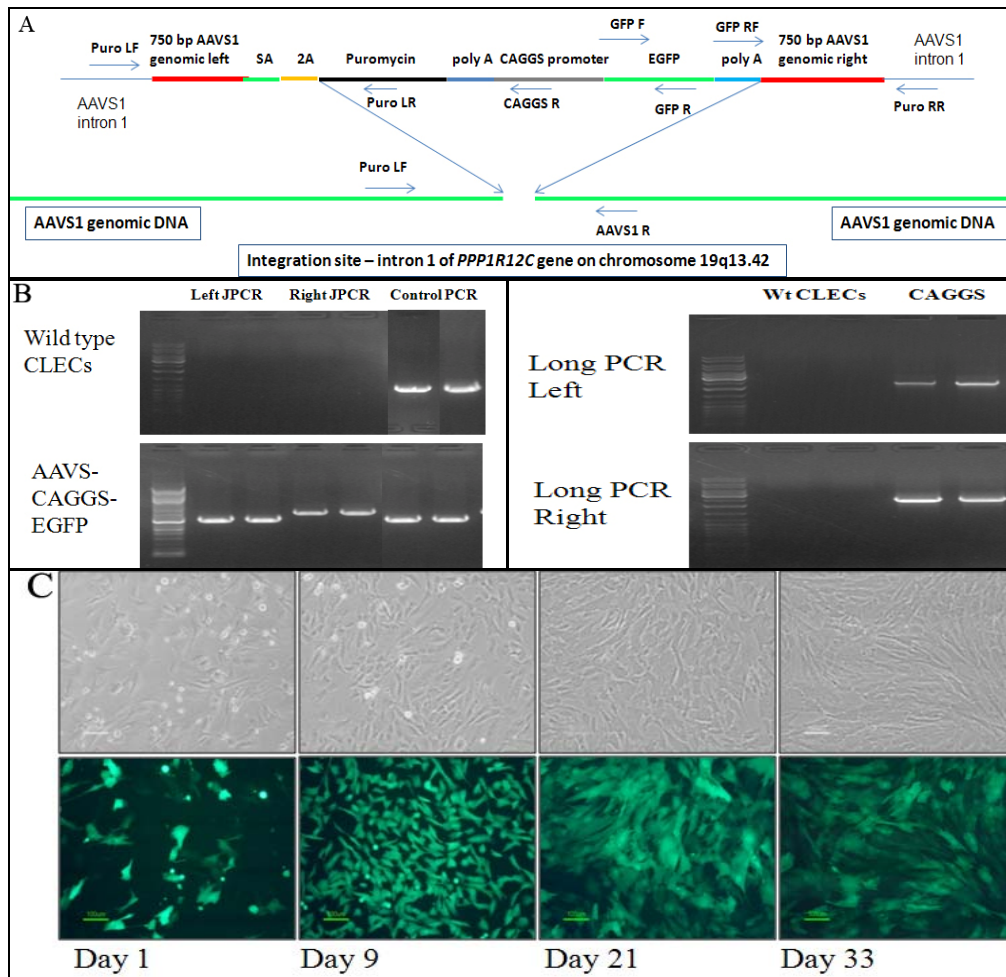
Figure 45 **Complete integration of AAV-CAGGS-GFP into AAVS1 locus of CLECs. (A)** Schematic showing homologous recombination-mediated integration of AAV-CAGGS EGFP into AAVS1 locus and the primer sets (left junction: Puro LF; Puro LR; right junction: GFP RF; Puro RR) used for integration junction and overlapping PCR (Long PCR Left: Puro LF; CAGGS R, Long PCR Right: GFP F; Puro RR) analysis of site-specific integration. **(B)** CLEC#33 co-electroporated with AAV-CAGGS EGFP plasmid DNA (4.2-kb fragment consisting of promoterless puromycin resistance cDNA and CAGGS promoter -EGFP cDNA) and a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were selected with puromycin. Genomic DNA extracted from the resulting stable cell populations were analyzed by (**left**) integration junction PCR and (**Right**) long PCR for evidence of donor DNA integration at the AAVS1 locus. Left (Left JPCR amplicon; 1 kb) and right (Right JPCR amplicon; 1.3 kb) integration junctions were amplified with vector specific and genome specific primers as indicated above, while control PCR was performed with genome specific primers amplifying a 900-bp region of the AAVS1 locus (all done in duplicate). Two different long PCR encompassing most of the integrated transgene (Long PCR Left; 2.3 kb; Long PCR Right, 2.5 kb) were performed with vector specific and genome specific primers as indicated above. Similar PCR reactions performed on unmodified wild type CLECs served as controls showing lack of donor integration. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Identity of junction PCR products and long PCR products were confirmed by sequencing. **(C)** Brightfield and fluorescence images of CLECs on day 1 post-electroporation, and on days 9, 21 and 33 (i.e. on completion of puromycin selection) with AAV-CAGGS EGFP plasmid DNA.

2.3.3.3.   AAVS1 site-specific integration of complete 9-kb Hfer hybrid FVIII cassette in CLECs

Having demonstrated the capacity to target transgene integration into the AAVS1 locus and to enrich for cells with such integration using a promoter trap strategy, we proceeded to evaluate if complete integration of the 9-kb hybrid FVIII cassette was possible and whether site-specific transgene integration at the AAVS1 locus would support durable FVIII expression.

Integration junction PCR and long PCR done on genomic DNA extracted from puromycin-resistant cells derived following co-electroporation with pSA-2A-puro-Hybrid FVIII and a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs provided evidence for site-specific integration of the entire 9-kb hybrid FVIII transgene cassette at the AAVS1 locus (**Figure 46B**). Sequencing of the two long PCR products encompassing the AAVS1 genomic region and the integrated vector confirmed integration of the entire transgene cassette without loss of any vector or genomic DNA sequences at the integration junctions or within the integrated vector (**Appendix 6**).

Cells electroporated with pSA-2A-puro-hybrid FVIII with or without ZFNs (indicated as "Puro FVIII ZFN" and "Puro FVIII", respectively in **Figure 46C**) but not puromycin-selected were incapable of durable FVIII expression as evident by undetectable levels of FVIII at day 37 post-electroporation. This was consistent with loss of episomal pSA-2A-puro-Hybrid FVIII vector in the absence of transgene integration ("Puro FVIII") and failure to enrich for cells with transgene integration in the absence of puromycin selection ("Puro FVIII ZFN"). On the other hand, durable FVIII expression was observed for at least 37 days post-electroporation in puromycin-resistant cells derived from electroporating pSA-2A-puro-Hybrid FVIII and a plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs ("Puro FVIII ZFN selected"). Puromycin-resistant cells had FVIII expression levels of 2131 ± 17.47 mUnits/ 10^6 cells/ 24 hr at day 37 post-electroporation compared to initial FVIII levels of 952.1± 8.365 mUnits/ 10^6 cells/ 24 hr from unselected cells at 1 day post-electroporation, showing durable FVIII expression after puromycin selection enriched for FVIII-expressing cells (**Figure 46C**). These data collectively showed that transgene integration at the AAVS1 locus could support durable transgene expression and corroborated with our earlier findings of durable EGFP expression following transgene integration at this locus (**Figure 45C**) as well as other reports that the AAVS1 locus is permissive for durable expression of integrated transgenes[8-10].

In order to estimate the ratio of off-target to on-target integrations, we performed integration junction PCR and vector-specific PCR on genomic DNA extracted from CLECs stably integrated with Puro FVIII vector following ZFN treatment (Puro FVIII ZFN selected cells). PCR products were quantified by densitometry. The band intensities of vector PCR products (vector PCR FVIII A3 domain and vector PCR FVIII C1 domain), reflecting on-target and off-target integrations were expressed as a ratio of integration junction PCR products, reflecting on-target integration (Left integration junction). A ratio close to 1.0 would indicate mainly on-target integration while ratios below or greater than 1.0 might suggest vector deletion or significant off-target integrations, respectively. Our analysis determined that the ratio of vector PCR FVIII A3 domain: integration junction PCR was 0.91 and that of vector PCR FVIII C1 domain: integration junction PCR to be 1.16 (**Figure 46D**). Given that both ratios were close to 1.0, it could be inferred that off-target integrations or vector deletions were not major features in these cells, although neither could be definitively excluded from these data.

The proliferative capacity of naive ZFN-untreated CLECs, CLECs electroporated with pSA-2A puro-hybrid FVIII cDNA only and stable CLECs derived from co-electroporation of pSA-2A-puro-hybrid FVIII and a bicistronic plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were evaluated by *in vitro* colony formation assay and MTS assay. As assessed by the MTS assay (**Figure 46F**) proliferation of untreated CLECs (O.D 405 nm = 0.563 ± 0.027) and stable CLECs with Puro FVIII integration (O.D 405 nm = 0.484 ± 0.03) was not statistically different ($P = 0.094$). Comparatively, CLECs transiently electroporated with pSA-2A-puro-hybrid FVIII cDNA only (O.D 405 nm = 0.723 ± 0.027) were significantly more proliferative (state *P* value here). Data from *in vitro* colony formation assay (**Figure 46E**) corroborated MTS assay data. There was no significant difference ($P = 0.387$) in the number of colonies formed from initial seeding of 100 naive untreated CLECs (Wt CLECs = 6.0 ± 0.6) or stable CLECs integrated with pSA-2A-puro-hybrid FVIII (Puro FVIII ZFN = 8.7 ± 2.2) but significantly more colonies formed from CLECs transiently electroporated with Puro FVIII only (Puro FVIII = 11.0 ± 1.0) (state *P* value here). In summary, data from both MTS and *in vitro* colony formation assays indicated comparable growth rates of stable CLECs with genomic integration pSA-2A-puro hybrid FVIII and untreated naive CLECs.
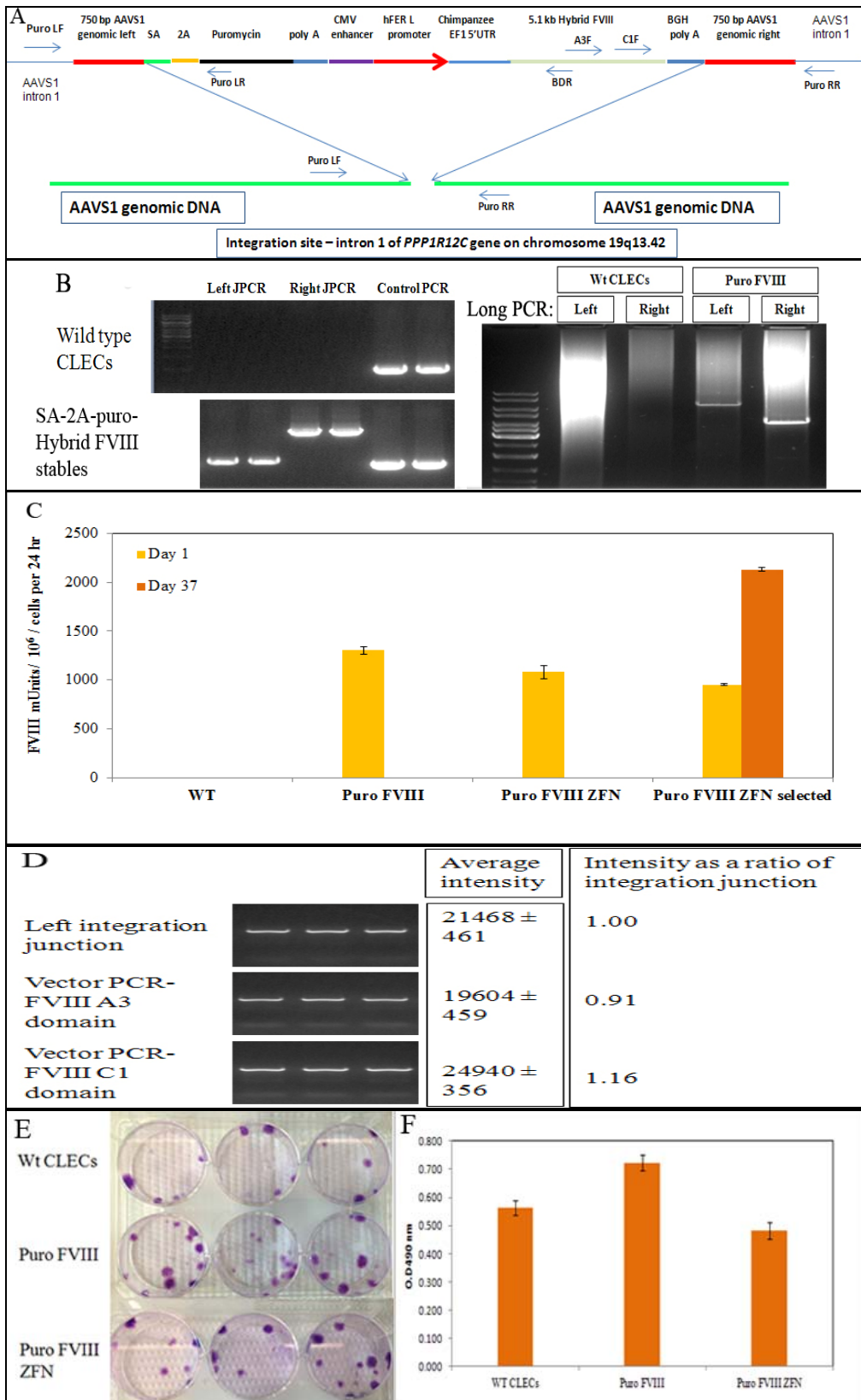
Figure 46   Complete integration of pSA-2A-puro-hybrid FVIII into AAVS1 locus of CLECs. (**A**) Schematic showing homologous recombination-mediated integration of pSA-2A-puro-hybrid FVIII into AAVS1 locus and the primer sets (left junction: Puro LF, Puro LR; right junction: C1F, Puro RR) used for integration junction and overlapping PCR (Long PCR Left: Puro LF; BDR; Long PCR Right: A3F; Puro RR) analysis of site-specific integration. (**B**) CLEC#33 co-electroporated

with pSA-2A-puro-hybrid FVIII plasmid DNA (8.9-kb fragment consisting of promoterless puromycin resistance cDNA and human ferritin promoter-hybrid FVIII cDNA) and a plasmid encoding both left and right AAVS1 enhanced Sharkey ZFNs were selected with puromycin. Genomic DNA extracted from the resulting stable cell populations were analyzed by (**left**) integration junction PCR and (**right**) long PCR for evidence of donor DNA integration at the AAVS1 locus. Left (Left JPCR amplicon, 1 kb) and right (Right JPCR amplicon, 3.187 kb) integration junctions were amplified with vector specific and genome specific primers as indicated above, while control PCR amplified a 900-bp region of the AAVS1 locus (all done in duplicate). Two different long PCR encompassing the entire integrated transgene (Long PCR Left, 6.859 kb; Long PCR Right, 4.156 kb) were performed with vector specific and genome specific primers as indicated above. Identical PCR performed on unmodified wild type CLECs served as controls for absence of donor integration. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and QuantityOne software. Identity of junction PCR products and long PCR products were confirmed by sequencing. (**C**) Conditioned media of CLECs that were either unelectroporated (WT) or electroporated with pSA-2A-Puro-hybrid FVIII (Puro FVIII), pSA-2A-Puro-hybrid FVIII and a plasmid encoding left and right AAVS1 enhanced Sharkey ZFNs with (Puro FVIII ZFN selected) or without (Puro FVIII ZFN) a 7-day puromycin selection starting 1 week after electroporation were assayed for FVIII activity on days 1 and day 37 post-electroporation (Coamatic® FVIII assay, Chromogenix). Data are mean ± SEM; n=3 per group. (**D**) CLECs integrated with Puro hybrid FVIII cDNA were evaluated for the ratio of on-target to off-target integrations by a genomic PCR method. Band intensities by densitometry of PCR products from integration junction PCR (on-target integration events) and vector PCR (combination of on-target and off-target integration events) were expressed as a ratio of vector specific PCR product: integration junction PCR product. (**E**) Crystal violet staining of cell colonies 14 days after initial seeding (in triplicate wells) of 100 (**top; Wt CLECs**) naive untreated CLECs, (**middle; Puro FVIII**) CLECs electroporated with Puro hybrid FVIII cDNA only or (**bottom; Puro FVIII ZFN**) stable CLECs derived from co-electroporation of Puro hybrid FVIII and ZFN. Images of culture dishes containing crystal violet stained cells. For each seeding density, the number of colonies were enumerated and expressed as mean colony counts ± standard error of triplicates. (**F**) Naive untreated CLECs (**Wt CLECs**), CLECs electroporated with Puro hybrid FVIII cDNA only (**Puro FVIII**) or stable CLECs derived from co-electroporation of Puro hybrid FVIII and ZFN (**Puro FVIII ZFN**) were seeded at an initial seeding density of 100 cells per well (96 well plate) and quantified 7 days later for cell proliferation using the MTS assay (CellTiter 96®Aqueous One solution cell proliferation assay kit), Data are mean absorbance readings ± SEM; n=6 per group.

.

2.3.3.4.   Effect of transgene integration at AAVS1 locus on endogenous *PPP1R12C* transcription

Insertion of splice acceptor (SA) sequences into the genome by integrated vectors may potentially alter endogenous gene expression. To investigate if this occurred at the AAVS1 locus, endogenous *PPP1R12C* transcripts were assessed in unmodified CLECs (Wt CLECs) and CLECs with stable ZFN-mediated integration of donor vectors (AAVS1 SA-2A-puro-pA donor, AAVS-CAGGS-EGFP and pSA-2A-puro-hybrid FVIII) by RT-PCR for changes in expression levels.

The mean *GAPDH*-normalized *PPP1R12C* transcript levels in stable CLECs integrated with AAVS1 SA-2A-puro-pA donor, AAVS-CAGGS-EGFP and pSA-2A-puro-Hybrid FVIII were determined to be 0.005 ± 0.0006, 0.461 ± 0.39 and 0.443 ± 0.03, respectively, of the normalized *PPP1R12C* transcripts levels of control CLECs (Wt CLECs) (**Figure 47**). This translates to a 200-fold (AAVS1 SA-2A-puro-pA donor) and approximately 2-fold (AAVS-CAGGS-EGFP and pSA-2A-puro-Hybrid FVIII) decrease in *PPP1R12C* transcript levels as compared to control CLECs, consistent with bi-allelic and mono-allelic transgene integrations in the respective stable cultures ($P<0.05$). These data were concordant with genomic PCR data (**Figure 44B**; "Long PCR") which showed strong amplification of the integrated 1-kb puromycin cassette and much weaker amplification of the endogenous locus, consistent with predominantly bi-allelic integration of AAVS1 SA-2A-puro-pA donor in this population. Other investigators have indeed reported similar declines in *PPP1R12C* transcription following integration of donor vectors with splice acceptor (SA) sequences into AAVS1 locus[9, 11]. Although no adverse consequences have been reported following complete disruption of the endogenous *PPP1R12C* gene, it may be useful and interesting to determine if biologically meaningful global transcriptional changes arise when endogenous *PPP1R12C* gene expression is extinguished.
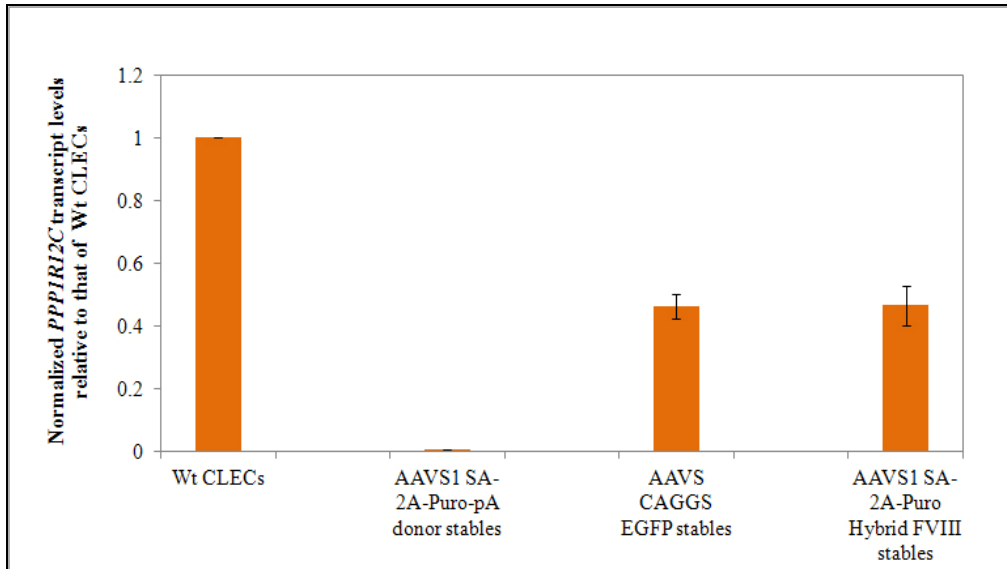
Figure 47 **Endogenous *PPP1R12C* transcription in transgene-integrated CLECs.** Quantitative RT-PCR was performed on control untreated CLECs (Wt CLECs) and puromycin-resistant stable CLECs, derived from co-electroporation with AAVS1 SA-2A-puro-pA donor, AAVS-CAGGS-EGFP or pSA-2A-puro-hybrid FVIII and a plasmid encoding for AAVS1 enhanced Sharkey ZFNs to assay *PPP1R12C* (exon 4 -6) and *GAPDH* transcript levels. *GAPDH-* normalized *PPP1R12C* expression levels (as determined by the $2^{-\Delta\Delta Ct}$ method) are shown relative to that of control CLECs (Wt CLECs). Data are mean ± SEM, n=3 experiments per group and 3 replicates per sample.

2.3.3.5.   Transcriptome analysis of CLECs with transgene integration at AAVS1 locus

The transcriptome of naive unmodified CLECs and a bulk population of puromycin-resistant, ZFN-treated CLECs with targeted integration of puro hybrid FVIII cassette were analyzed on Affymetrix HU133 plus 2.0 arrays. Genes whose expression differed by greater than 2-fold compared to unmodified CLECs were analyzed further.

Targeted transgene integration into intron 1 of the *PPP1R12C* gene has the capacity to perturb the expression of the endogenous gene[8, 9, 11] and may even affect the expression of its potential interacting partners, other protein phosphatases within the same family, potential downstream effector genes[12-15] and neighboring genes within the vicinity of transgene integration. We thus focused analysis of the transcriptome data to a close scrutiny of the endogenous *PPP1R12C* transcript and other genes of the protein phosphatases family (n=75), potential interacting partners of *PPP1R12C* predicted by Gene Network Central™ (http://www.sabiosciences.com/genenetwork/genenetworkcentral.php) and String 9.0 (http://string-db.org/) (n=24), potential downstream effector genes (n =63) and neighboring genes within a 1 Mb window of transgene integration site (n=45). Selected genes were evaluated by quantitative RT-PCR to verify transcriptome data.

Concordant with RT-PCR (**Figure 47**), transcriptome data also showed a 2.2-fold decline in endogenous *PPP1R12C* expression. The insertion of a new splice-acceptor sequence into intron 1would very likely have truncated the endogenous *PPP1R12C* transcript, thus reducing its level. Given the 2-fold reduction in *PPP1R12C* transcription, it would be reasonable to postulate that the majority of cells had attained monoallelic rather than biallelic integrations, since the latter would more likely completely extinguish *PPP1R12C* expression. It is worth noting that as ES/iPS cells tolerate biallelic disruption of *PPP1R12C* [1, 10], there appears to be functional redundancy of its role in cells . The normal morphology and growth characteristics of genome-modified CLECs despite a 50% reduction in *PPP1R12C* expression further supports a less than crucial role of *PPP1R12C* for cell survival and growth, since partial or even complete disruption of this gene is also tolerated by CLECs.

We next examined the effects of *PPP1R12C* down-regulation on its potential interacting partners as predicted by Gene Network Central™ and String 9.05. A total of 24 genes were analyzed, of which only 4 showed significant changes in expression in genome-modified CLECs compared to control CLECs [Up-regulated: *CDC*6 (2-fold), *DUSP*6 (5.5-fold) ; Down-regulated: *DUSP*16 (5.3-fold), *DUSP*1 (2.4-fold)].

However, quantitative RT-PCR could verify a significant change in only one of these 4 genes, *DUSP*6, which was up-regulated by 4.2–fold in genome-modified CLECs compared to control CLECs (**Figure 48**). Over-expression of Dual Phosphatase 6, *DUSP*6, has been associated with reduction of cellular proliferation and induction of apoptosis[16] in a lung cancer cell line and would be expected to have tumor suppressive roles rather than oncogenic effects. No significant changes in transcript levels were detected by quantitative RT-PCR for the remaining 3 genes.

As *PPP1R12C* is a regulatory subunit of myosin-binding protein phosphatase 1 delta, we further evaluated all potential myosin-related transcripts (n=31) as well as predicted downstream effector genes (n=32) based on limited publications on *PPP1R12C* [12-15]. Of 63 potential genes analyzed, 7 had significant changes in genome-modified CLECs compared to control CLECs [Up-regulated: *MYL12A* (3.2-fold), *MYH*15 (14-fold); Down-regulated: *CDC*42 (2.6-fold), *CDC42EP4* (2.3-fold), *MYH8* (7.7-fold), *MYLIP* (3.6-fold)]. Altered expressions of these significantly altered genes are not known to be associated with any known disease phenotype, by reference to the OMIM database. We cannot exclude the possibility that at least some of the transcriptional alterations observed directly or indirectly resulted from *PPP1R12C* down-regulation. This question could be resolved by complete knock-down studies of *PPP1R12C* expression.

Another key concern of integrating vectors is the potential to perturb the expression of genes neighboring the integration site. We therefore analyzed our transcriptome data for perturbation of genes located within a 1 Mb window centered around the integration site in intron 1 of *PPP1R12C*. Of 43 genes in this window, only 7 were expressed in CLECs and none were altered in expression. Quantitative RT-PCR results corroborated the transcriptome data for all 7 genes (**Figure 48**). This finding was not unexpected, given that our integrating vector did not contain any enhancer elements and the presence of natural insulator elements at the AAVS1 locus[17].

We further extended our analysis to 75 genes of the protein phosphatase family (PPP1 to PPP6). Analysis of the transcriptome data revealed that 5 genes within this family showed changes in gene expression in genome-modified CLECs compared to control unmodified CLECs [Up-regulated: *PPP2R2A* (2.2-fold), *PPP4R2* (3-fold), *PPP4R4* (2.7-fold); Down-regulated: *PPP2R5A* (2.7-fold), *PPP1R36* (2.7-fold)]. It is highly speculative as to whether changes in other members of the large protein phosphatase family were direct effects of or compensatory responses to *PPP1R12C* down-regulation or were completely unrelated.

Lastly, all the above-mentioned genes with significant changes in expression were mapped to a database of potential oncogenes. This revealed that none of the altered genes are potential oncogenes.

At present, the *PPP1R12C* gene has not been well-studied, and its functions and downstream effects are thus poorly understood. While we could not confidently identify if any functional pathways were directly affected as a result of *PPP1R12C* down-regulation from transgene integration, perturbations in gene expression were noted in a limited number of transcripts directly or indirectly associated with *PPP1R12C*. The functional effects of changes to these limited number of genes remains to be evaluated, although they exerted no effect on cell growth or morphology. **Appendix 9** is a complete list of all genes evaluated in this transcriptome study.
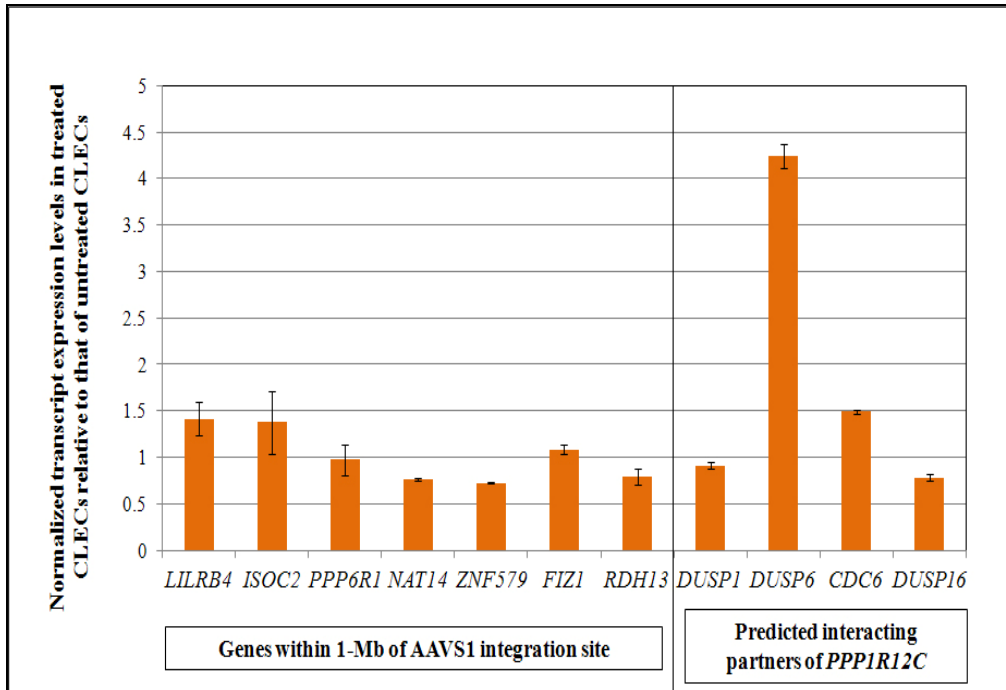
Figure 48 **RT-PCR to verify transcriptional changes to selected genes identified by genome-wide transcriptome profiling.** Quantitative RT-PCR was performed on control untreated CLECs and a bulk population of CLECs with ZFN-mediated integration of Puro Hybrid FVIII transgene at the AAVS1 locus. Data show transcript levels of neighboring genes residing within 1-Mb of the AAVS1 integration site, predicted interacting partners of *PPP1R12C* which were deemed to be significantly altered by transcriptome analysis and a housekeeping gene, *GAPDH*. *GAPDH*-normalized transcript expression levels (as determined by the $2^{-\Delta\Delta Ct}$ method) in treated CLECs are shown relative to that of control untreated CLECs. Data are mean ± SEM, n=2 experiments per group and 3 replicates per transcript.

2.3.3.6.   Deep-sequencing of top-10 potential off-target sites for AAVS1 ZFNs

Although ZFNs are designed to be highly specific and to target a unique site in the genome, the possibility for off-target binding and activity elsewhere in the genome cannot be excluded. There remains a rare possibility that ZFNs pairs may bind as homo- or heterodimers and cleave at genomic regions having similar sequence as the intended target recognition sites. Such off-target effects would be anticipated to create new indels following NHEJ repair. We thus evaluated the precision of AAVS1 ZFN-mediated transgene insertion in CLECs by massively parallel sequencing of the 10 most likely off-target sites for AAVS1 ZFNs bioinformatically predicted by SELEX analysis[1] (**Appendix 10**).

The endogenous AAVS1 locus and aforementioned top-10 potential off-target sites were amplified from genomic DNA of naive untreated CLECs and a bulk population of puromycin-resistant ZFN-treated CLECs with targeted integration of Puro FVIII vector. PCR amplicons were sequenced using Illumina MiSeq to perform 150-bp paired-end deep sequencing. Sequence data were analyzed for indels at potential ZFN recognition sites within each amplicon. PCR amplicons specific to the AAVS1 locus targeted by AAVS1 ZFNs served as positive controls for indel detection. In order to determine the sensitivity of the deep sequencing platform for detecting indels, the unmodified AAVS1 amplicon was spiked with a synthetic mutant AAVS1 amplicon having a 5-bp deletion at the AAVS1 ZFN target site at ratios of 1:10 (10% mutant), 1:100 (1% mutant), 1:500 (0.2% mutant) and 1:1000 (0.1% mutant).

Targeted deep sequencing consistently detected indels at all concentrations of the spike-in controls (0.1% - 10% mutants) and gave a highly positive linear correlation ($R_2 = 0.999$) between the actual percentage of spike-in controls and experimentally detected percentage of indels (**Appendix 11**). The spike-in-control experiments established the feasibility of targeted deep-sequencing to detect mutants (5-bp deletions) in as few as 0.1% of the population. Thus it was reasonable to accept that rare off-target events (manifest as indels) present in even 0.1% of a mixed population of ZFN-treated stable CLECs with puro FVIII transgene integration would be detected by targeted deep-sequencing.

Our analysis of the top-10 potential off-target sites (OT1 – OT10) in CLECs showed that the AAVS1 ZFNs did indeed induce indels at low frequencies in only 1 of the 10 potential off-target sites. Two different types of indels, viz. a 4-bp deletion (UCSC position 141507040) and a 1-bpdeletion (UCSC position 141507072), were

found at OT1, an intergenic region in chromosome 8q24.3, at frequencies listed in **Table 3**. It is possible that indels common to both untreated CLECs and CLECs with Puro FVIII cDNA integration were technical artifacts introduced either during PCR generation and/or sequencing of amplicons. Aside from indels detected in OT1, no other indels specific to CLECs with Puro FVIII cDNA integration were detected at significant levels. Low frequency indels at OT6 were common to both untreated CLECs and ZFN-treated CLECs and thus could not be attributed to the off-target. Evaluation of our transcriptome data failed to reveal any change in transcript levels between untreated CLECs and ZFN-treated CLECs for off-target intronic sites (OT3, OT6, OT7, OT8, OT9 and OT10).

In summary, targeted deep sequencing appears to be a useful and sensitive tool for evaluating the precision of ZFNs designed for locus specificity. The use of AAVS1 ZFNs in CLECs resulted in low frequency indels at OT1. However, as OT1 is intergenic and substantially distant from protein-coding genes, low frequency cleavage and repair at this off-target site may be functionally silent.

Table 3   **Deep sequencing of top-10 potential off-target sites**

Site	Chromosomal locus	UCSC chromosomal location	Untreated CLECs			Puro FVIII ZFN stable CLECs			Transcriptional change
			Indels (frequency)	Total reads	% Indels	Indels (frequency)	Total reads	% Indels	
AAVS1	Chr 19 Intron 1 of *PPP1R12C*	55627135 55627136 55627114	None detected	92804	0	1 bp del (2285) 11 bp del (298) 41 bp del (83)	34974 35037 33195	6.53 0.85 0.25	
OT1	Chr 8 Intergenic	141507040 141507072 141507072 141507045 141507072 141507072	1 bp del (883) 2 bp del (32) 1 bp insertion (60) 2 bp insertion (15)	11707 11707 11707 11707	7.54 0.27 2.83 0.13	4 bp del (143) 1 bp del (532) 1 bp del (143) 1 bp insertion (212)	10500 8393 10434 8393	1.36 6.34 1.37 2.53	
OT2	Chr 10 Intergenic		None detected	82178	0	None detected	93049	0	
OT3	Chr 4 Intron 1 of *RGS12*		None detected	49810	0	None detected	67083	0	Not expressed
OT4	Chr 10 Intergenic		None detected	104186	0	None detected	110320	0	
OT5	Chr 9 Intergenic	138563405	1 bp del (84)	53278	0.16	None detected	94464	0	
OT6	Chr 14 Intron 1 of *BEGAIN1*	101033142	1 bp del (72)	40570	0.18	1 bp del (65)	45737	0.14	Not expressed
OT7	Chr 7 Intron 14 of		None detected	108022	0	None detected	107522	0	No change

	GRB10							
OT8	Chr 16 Intron 2 of *NPIPL1*/ Exon 1 of *LAT*		None detected	89961	0	None detected	130644	0	*NPIPL1* (Not expressed) *LAT* (Not expressed)
OT9	Chr 19 Intron 7 of *STK11*	1224724	1 bp del (74)	57423	0.13	None detected	97637	0	No change
OT10	Chr 12 Intron 4 of *FAIM2*		None detected	97602	0	None detected	86660	0	Not expressed

The AAVS1 ZFN binding site (AAVS1) and the top-10 potential off-target sites (OT1 to OT10) of untreated CLECs and ZFN-treated CLECs with stable integration of Puro hybrid FVIII cassette (Puro FVIII ZFN stable CLECs) were evaluated by targeted deep-sequencing. Chromosomal locus of the target sites investigated and corresponding USCS chromosomal location numbers, total mappable reads analyzed, the types and frequencies (given in parenthesis) of indels detected and their corresponding percentages are summarized. Where off-target sites occurred within a protein-coding gene, effects on the corresponding genes were evaluated by reference to the transcriptome datasets for transcriptional changes. For off-target sites without any significant indels and where no UCSC chromosomal numbers are reported, total reads refer to the sum of all mapped sequencing reads analyzed for that particular locus. Where UCSC chromosomal locations are reported, the total reads refer to mapped sequencing reads specific to that particular chromosomal location.

2.3.3.7.   Spectral karyotyping of stable CLECs with Puro hybrid FVIII cassette integration at the AAVS1 locus

ZFN-treated CLECs with stable integration of Puro hybrid FVIII transgene cassette at the AAVS1 locus, were investigated for chromosomal rearrangements by spectral karyotype analysis. Evaluation of 23 metaphases revealed normal karyotype (**Figure 49**) and no evidence of structurally abnormal chromosomes within the resolution of this technique[18]. These data suggested that ZFN treatment did not induce gross chromosomal abnormalities and that CLECs do not incur chromosomal anomalies during the course of puromycin selection or continuous *in vitro* culture for up to 1 month.

Figure 49 **Spectral karyotype of ZFN-treated CLECs.** A representative image of chromosomes from a single metaphase (out of 23 metaphases analyzed) showing 46 chromosomes, XX genotype and normal spectral karyotype.

2.3.4.   Evaluation of HSV-thymidine kinase as a suicide gene to eliminate cells with off-target integrations

In a previous section, we provided evidence for site-specific integration of a 9-kb hybrid FVIII donor DNA in G418-selected CLECs. However, G418 selection enriches not only for cells with stable AAVS1 site-specific integration but also for cells in which random integrations have occurred. We considered that incorporation of a HSV-thymidine kinase gene cassette into the donor vector in combination with gancyclovir selection could eliminate cells with off-target integrations. The HSV-TK cassette was cloned into the pZDonor hybrid FVIII construct outside the AAVS1 homology arms so that HSV-TK cassette would be retained only in the genome of cells where random integrations had occurred. Gancyclovir selection should kill cells expressing HSV-TK (random integration) while not affecting cells which do not express HSV-TK (site-specific integration).

2.3.4.1. Effect of gancyclovir elimination of TK007-expressing CLECs

We first compared the efficacy of gancyclovir selection to kill CLECs stably and randomly integrated with either wild type HSV-TK or a codon-optimized and improved HSV-TK007 cassette[19]. CLECs stably expressing HSV-TK or HSV-TK007 were first selected for G418 resistance, and then exposed to increasing concentrations of gancyclovir. The readout of this trial experiment was cell viability assessed by the MTS assay.

RT-PCR analysis (**Figure 50A**) of stable cell cultures derived from HSV-TK and HSV-TK007 electroporated samples confirmed expression of the HSV-TK and HSV-TK007 transcripts, respectively. PCR analysis of genomic DNA from the same stable cultures also showed genomic integration of both HSV-TK and HSV-TK007 expression vectors. Taken together, these results confirmed expression of the respective versions of HSV-TK in stable cell cultures before testing for gancyclovir sensitivity.

MTS analysis (**Figure 50B**) showed decreased viability of CLECs expressing HSV-TK or HSV-TK007 when exposed to increasing concentrations of gancyclovir. Naive CLECs which did not express HSV-TK were resistant to gancyclovir. Their viabilities were not significantly reduced even at 10 μM gancyclovir i.e. 92.18 ± 5.49 % compared to without gancyclovir treatment. In contrast, CLECs expressing HSV-TK and HSV-TK007 had significantly reduced cell viabilities ($P<0.05$) at gancyclovir concentrations ranging from 0.1 to 10 μM compared to pre-treatment viabilities (no gancyclovir). Comparison between HSV-TK and HSV-TK007 showed significantly

reduced ($P<0.05$) cell viabilities with the latter at gancyclovir concentrations of 1 to 10 μM. At 1 μM gancyclovir, CLECs expressing HSV-TK007 had a viability of 20.61 ± 1.52% which was significantly lower ($P = 0.0002$) than the viability of CLECs expressing HSV-TK (56.25 ± 3.98%). Increasing gancyclovir concentrations beyond 1 μM further reduced viabilities of CLECs expressing HSV-TK and HSV-TK007 but a significant decline in viability was also noted for naive CLECs. As such, the optimal gancyclovir concentration for eliminating CLECs expressing HSV-TK was taken as 1 μM. Furthermore, since HSV-TK007 was more effective in mediating cell death in the presence of gancyclovir, it was incorporated in the final version of FVIII donor construct, pZDonor hybrid FVIII TK007.
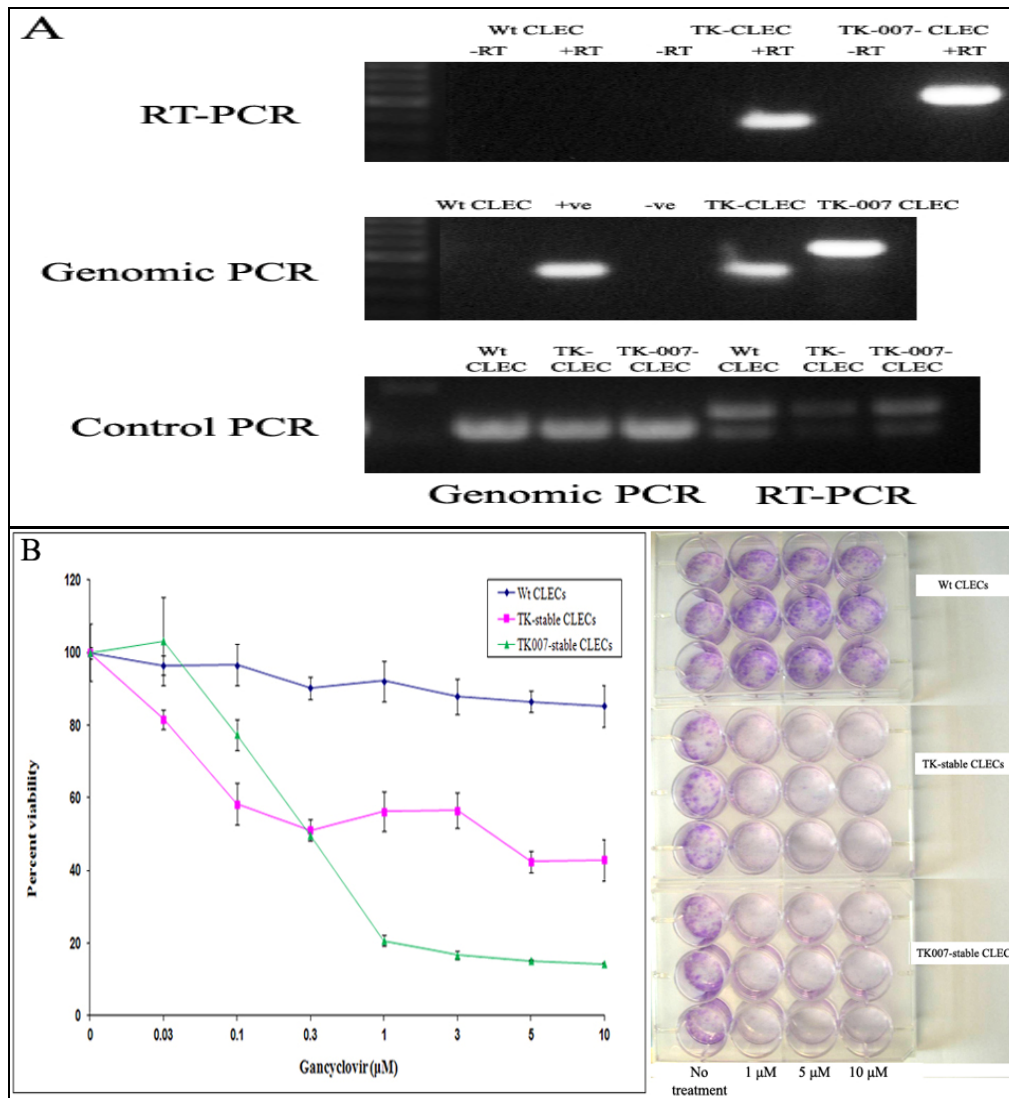
Figure 50 **Effectiveness of gancyclovir in eliminating HSV-TK-expressing CLECs.** (**A**) Samples from naive CLECs (Wt CLEC) and CLECs stably expressing either wild type HSV-TK (TK-CLEC) or codon-optimized HSV-TK007 (TK007-CLEC) were evaluated for expression of HSV-TK/TK007 transcripts and genomic integration of HSV-TK/TK007 cDNA by RT-PCR and genomic PCR, respectively, 1 month post-electroporation. "+RT" and "-RT" refer to 1st strand cDNA synthesis by reverse transcription reaction performed with or without reverse transcriptase, respectively. "+ve" Refers to amplification of HSV-TK cDNA cassette from a plasmid template while "-ve" refers to minus template amplification. Control PCR amplified a housekeeping gene, gamma-actin, from either genomic DNA (Genomic PCR) or reverse-transcribed 1st strand cDNA (RT-PCR) from the indicated samples. Amplification products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator. (**B**) Naive CLECs (Wt CLEC) and CLECs stably expressing either wild type HSV-TK (TK-CLECs) or codon-optimized HSV-TK007 (TK007-CLECs) were cultured at various concentrations of gancyclovir for 6 days and evaluated by MTS assay for viable cells. Data are mean ± SEM; $n = 4$ per concentration evaluated. The mean absorbance readouts for untreated cells from each group were used to assign a viability value of 100%. For each group, percent viability was the absorbance measurements of cells treated at various gancyclovir concentrations expressed as a percentage of the mean absorbance of untreated cells. Significantly greater cell killing was observed for TK007-CLECs (compared to TK-

251

CLECs) at gancyclovir concentrations of 1 to 10 μM (Student's *t*-test, p<0.05). **(C)** Crystal violet staining of naive CLECs (Wt CLEC) and CLECs stably expressing either wild type HSV-TK (TK-CLECs) or codon-optimized HSV-TK007 (TK007-CLECs) after 6 days of culture in normal culture media (No treatment) or in culture media supplemented with the indicated gancyclovir concentrations show unaffected cell viability in naive CLECs but greatly reduced viability in CLECs expressing HSV-TK or TK007 selected with gancyclovir at the indicated concentrations.

We attempted to enrich for cells with HR-mediated site-specific integration of the pZDonor hybrid FVIII TK007 by negative selection of G418-resistant CLECs with gancyclovir. Although our initial results with CLECs stably expressing TK007 indicated that selection with 1 μM gancyclovir could effectively eliminate TK007 expressing cells, subsequent studies with CLECs integrated with pZDonor hybrid FVIII TK007 showed that gancyclovir concentrations had to be higher than 2.5 μM before effective cell killing was achieved (**Figure 51B**). Untreated CLECs were unaffected by gancyclovir concentrations up to 100 μM (**Figure 51A**). Although gancyclovir selection eliminated CLECs expressing TK007, we repeatedly found that cells surviving the selection were non-proliferative and in a senescence-like state (**Figure 51D**). This senescence-like phenotype was not observed in G418-resistant CLECs prior to gancyclovir selection (**Figure 51C**). With this unexpected effect of high gancyclovir, we were unable to determine if incorporation of a suicide gene was effective in enriching for cells with HR-mediated site-specific transgene integration. All subsequent ZFN studies were therefore based on pZDonor hybrid FVIII construct rather than the pZDonor hybrid FVIII TK007 construct. Further work will be required to uncover the basis of quiescence and senescence induced by high gancyclovir concentration.
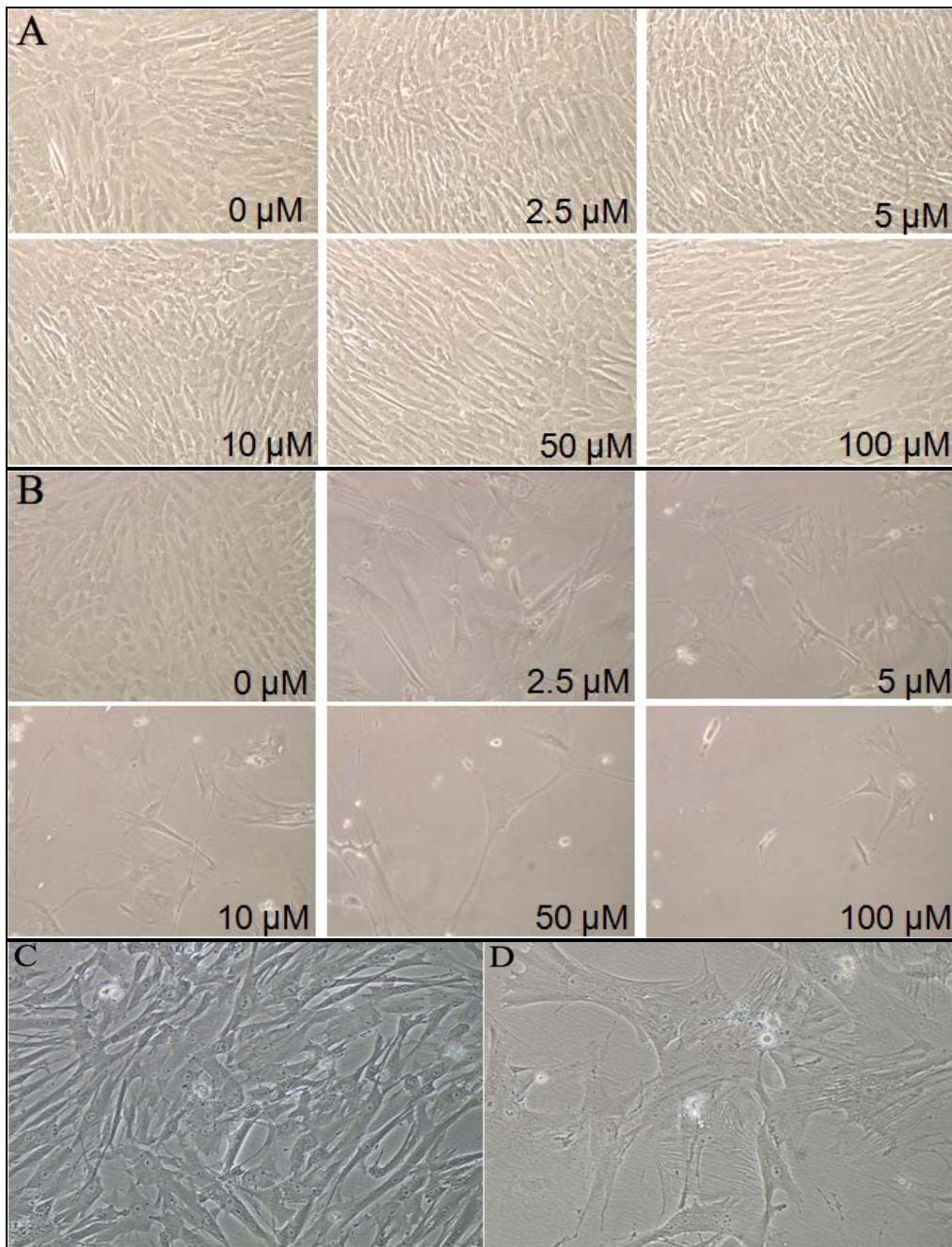
Figure 51   Gancyclovir selection of untreated and pZDonor hybrid FVIII TK007 construct integrated CLECs. Brightfield images of (**A**) untreated CLECs and (**B**) CLECs stably expressing MC1-HSV-TK007 following a 7-day selection with gancyclovir at the indicated concentrations shows effective killing of CLECs expressing HSV-TK007 at concentrations ≥ 2.5 μM. Brightfield images of G418-selected CLECs treated with pZDonor hybridF8-TK007 and ZFN (**C**) before gancyclovir selection show healthy and proliferative cells while (**D**) a senescence-like appearance is noted in quiescent cells that survived a 7-day selection with 100 μM gancyclovir.

2.3.5. References

1. Hockemeyer D, Soldner F, Beard C, Gao Q*, et al*. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. (2009) 27:851-857.

2. Moehle EA, Moehle EA, Rock JM, Rock JM*, et al*. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. Proc Natl Acad Sci U S A. (2007) 104:3055-3060.

3. Miller JC, Holmes MC, Wang J, Guschin DY*, et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. (2007) 25:778-785.

4. Guo J, Gaj T, Barbas CF. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol. (2010) 400:96-107.

5. Doyon Y, Vo TD, Mendel MC, Greenberg SG*, et al*. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. (2011) 8:74-79.

6. Doyon Y, Choi VM, Xia DF, Vo TD*, et al*. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat Methods. (2010) 7:459-460.

7. Rogakou EP, Pilch DR, Orr AH, Ivanova VS*, et al*. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem. (1998) 273:5858-5868.

8. Zou J, Sweeney CL, Chou BK, Choi U*, et al*. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood. (2011) 117:5561-5572.

9. Chang CJ, Bouhassira EE. Zinc-finger nuclease-mediated correction of α-thalassemia in iPS cells. Blood. (2012) 120:3906-3914.

10. Smith JR, Maguire S, Davis LA, Alexander M*, et al*. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. Stem Cells. (2008) 26:496-504.

11. Lombardo A, Cesana D, Genovese P, Di Stefano B*, et al*. Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat Methods. (2011) 8:861-869.

12. Hartshorne DJ, Ito M, Erdödi F. Myosin light chain phosphatase: subunit composition, interactions and regulation. J Muscle Res Cell Motil. (1998) 19:325-341.

13. Tan I, Ng CH, Lim L, Leung T*, et al*. Phosphorylation of a novel myosin binding subunit of protein phosphatase 1 reveals a conserved mechanism in the regulation of actin cytoskeleton. J Biol Chem. (2001) 276:21209-21216.

14. Surks HK, Richards CT, Mendelsohn ME. Myosin phosphatase-Rho interacting protein. A new member of the myosin phosphatase complex that directly binds RhoA. J Biol Chem. (2003) 278:51484-51493.

15. Mulder J, Ariaens A, van den Boomen D, Moolenaar WH*, et al*. p116Rip targets myosin phosphatase to the actin cytoskeleton and is essential for RhoA/ROCK-regulated neuritogenesis. Mol Biol Cell. (2004) 15:5516-5527.

16. Zhang Z, Kobayashi S, Borczuk AC, Leidner RS*, et al*. Dual specificity phosphatase 6 (DUSP6) is an ETS-regulated negative feedback mediator of oncogenic ERK signaling in lung cancer cells. Carcinogenesis. (2010) 31:577-586.

17. Ogata T, Kozuka T, Kanda T. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. J Virol. (2003) 77:9000-9007.

18. Fan YS, Siu VM, Jung JH, Xu J*, et al*. Sensitivity of multiple color spectral karyotyping in detecting small interchromosomal rearrangements. Genet Test. (2000) 4:9-14.

19. Preuss E, Treschow A, Newrzela S, Brücher D*, et al*. TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. Hum Gene Ther. (2010) 21:929-941.

2.4. Evaluation of other primary cell types

Whilst CLECs could potentially be developed for autologous cell therapy of pediatric patients, developing autologous cell therapy for most adults would other patient-derived primary cell types. Here we present a preliminary evaluation of primary human bone marrow-derived stromal cells (BMSCs), adipose-derived stromal cells (ADSCs) and dermal fibroblasts for their amenability for targeted gene integration at the AAVS1 locus, proper FVIII processing and secretion. These alternative cell types were chosen for their ease of isolation and culture, as well as their capacity to be expanded *in vitro* to clinically meaningful numbers.

2.4.1. Optimization of electroporation conditions for dermal fibroblasts, ADSCs and BMSCs

We first established electroporation conditions for efficient gene transfer to primary human dermal fibroblasts, ADSCs and BMSCs. Cells were electroporated with recommended Amaxa® solutions and an EGFP plasmid using different pre-set electroporation programs in an Amaxa® Nucleofector™ I. The transfection efficiency was the proportion of GFP-expressing cells determined by flow cytometry. Percent viability (where indicated) was calculated by trypan blue exclusion cell counts performed 24 hours post-electroporation.

We used primary human fibroblasts, commercially purchased primary human foreskin fibroblasts (Hs68), human patient-derived dermal fibroblasts KF1 and NF123 to optimize electroporation conditions. **Figure 52** summarizes gene transfer and viability (where available) data of human fibroblast cells. Hs68 was used in initial trials from which highest gene transfer efficiency was achieved with program U23 (32.1 % ± 0.40, Data are mean ± SEM, n=3). However, this program also resulted in lowest cell survival (25% viability). In order to recover as many electroporated cells as possible for downstream studies, all subsequent fibroblast electroporations were performed with program A24 which gave the best cell survival (66.5% viability) while having albeit with more modest transfection efficiency of 15.9 % ±0.03 (Data are mean ± SEM, n=3). It was of interest to note that transfection efficiencies of fibroblasts from different human donors such as KF1 (74.84% ± 0.161) and NF123 (24.43% ± 0.171) varied considerably although electroporated in identical conditions. This suggested that electroporation may need to be individually optimized. .

We used commercially purchased primary human bone marrow mesenchymal stem cells (Lonza MSC), human patient-derived bone marrow stromal cells (BMSC1, BMSC2 and Gan BM) and adipose-derived stromal cells (ADSC1) to establish

electroporation conditions from stromal cells. **Figure 53** summarizes gene transfer and viability (where available) data of cells electroporated with a Nucleofector™ I device using recommended programs. Lonza MSCs electroporated with the two recommended programs, C16 and U23, gave transfection efficiencies of 9.47% ± 1.154 and 39.53% ± 0.427 (Data are mean ± SEM, n=3) with survival rates of 46% and 33%, respectively. Subsequent electroporation of BMSC1, BMSC2 and ADSC1 were done using program U23 because it gave relatively high transfection efficiencies of 87.39% ± 0.242, 90.22% ± 0.06 and 76.54% ± 0.4, respectively. However, a gentler program, C17, was used to electroporate Gan BM cells to obtain transfection efficiency of 60.95 % ± 0.09, as they survived poorly with program U23. The variable transfection efficiencies among stromal cells from different human sources again pointed to the need to customize electroporation conditions to different cell types and from different donors.

Cell type/ treatment:	Hs68/ A24	Hs68/ U23	Hs68/ T16	Hs68/ V13	KF1/ A24	NF123/A24
Percent GFP +ve:	15.9	32.1	24.6	19.7	74.8	24.4
Percent viability:	66.5	25	46.7	46.7	ND	ND
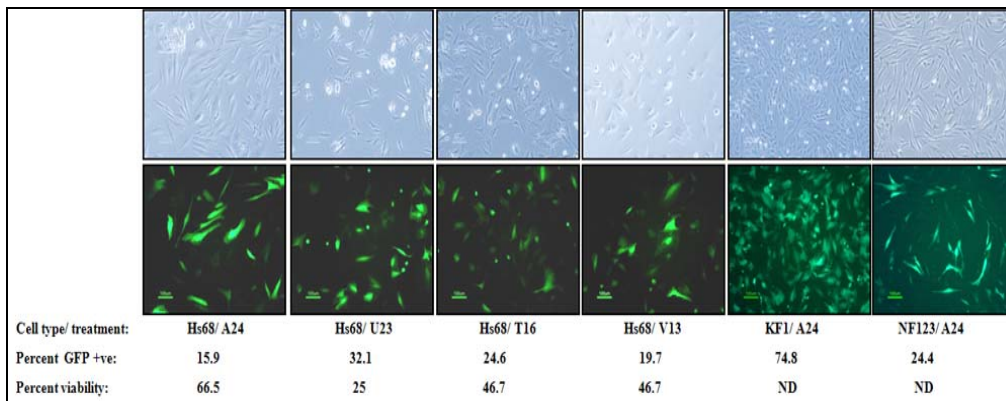
Figure 52   Optimization of electroporation conditions for primary human fibroblasts. Primary human foreskin fibroblasts (Hs86) and dermal fibroblasts (KF1 and NF123) were electroporated with 10 μg of EGFP reporter plasmid in an Amaxa® Nucleofector™ I device and the indicated electroporation programs. Representative brightfield and fluorescence images taken 1 day post-electroporation are shown (original magnification x100). Flow cytometric analysis for GFP expression (percent GFP +ve) and percent viability as determined by trypan blue exclusion cell counts were performed 1 day post-electroporation. (ND – not determined)

Cell type/ treatment:	Lonza MSC/ C16	Lonza MSC/ U23	BMSC1/ U23	BMSC2/ U23	ADSC1/ U23	Gan BMSC/ C17
Percent GFP +ve :	11	40	87.4	90.2	76.5	60.9
Percent viability:	46	33	ND	ND	ND	ND
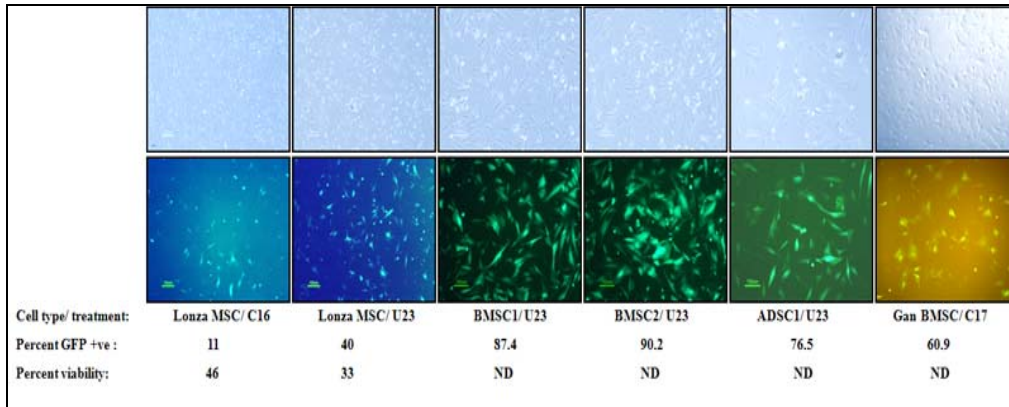
Figure 53   Optimization of electroporation conditions for primary human bone marrow- and adipose-derived stromal cells. Primary human bone marrow-derived stromal cells (Lonza MSC, BMSC1, BMSC2, Gan BM) and adipose-derived stromal cells (ADSC1) were electroporated with 10 μg of EGFP reporter plasmid with a Nucleofector™ I device and the indicated electroporation programs. Representative brightfield and fluorescence images taken 1 day post-electroporation are shown (original magnification x100). Flow cytometric analysis for GFP expression (percent GFP +ve) and percent viability as determined by trypan blue exclusion cell counts were performed 1 day post- electroporation. (ND – not determined)

2.4.2. Evidence of site-specific genomic cleavage in adult primary human cells

Having optimized electroporation conditions for BMSC, ADSC and dermal fibroblasts, we next investigated if these adult primary human cell types were amenable to ZFN mediated site-specific genome cleavage.

Two samples of each cell type (BMSC: BMSC 1,2; ADSC: ADSC 1,2; dermal fibroblast: NF123, KF1) were electroporated with plasmid encoding enhanced Sharkey ZFNs and incubated under mild hypothermic condition of 30°C for 2 days after an initial overnight incubation at 37°C. Genomic DNA harvested from electroporated cells 4 days post-electroporation was used for PCR amplification of the AAVS1 locus resulting in a 469-bp amplicon which was used for the *Cel*-1 nuclease assay for heteroduplex DNA as evidence of genomic cleavage and repair. The presence of indels at the AAVS1 site is indicated by *Cel*-1 nuclease mediated cleavage of the 469-bp amplicon into two fragments of 287 bp and 182 bp. Densitometry of modified (*Cel*-1 cleaved) and unmodified genomic locus (uncleaved) revealed targeted genome modification efficiencies of 26.15 ± 3.34 % for BMSC1, $22.38 \pm 0.79\%$ for BMSC2, $22.05 \pm 1.17\%$ for ADSC1, 19.69 ± 1.61 %, for ADSC2, $64.97 \pm 3.48\%$ for NF123 and $21.33 \pm 0.60\%$ for KF1 (**Figure 54**). Genome modification efficiencies did not appear to be correlated with transfection efficiency. Significantly higher levels of genome modification were observed for human fibroblast sample NF123 ($64.97 \pm 3.48\%$) compared to another fibroblast sample, KF1 ($21.33 \pm 0.60\%$), despite similar transfection efficiencies. Intrinsic differences specific to the cell samples may underlie differences in genome modification efficiencies achieved.

In conclusion, our data suggests that using ZFNs to induce site-specific genomic cleavage and promote homologous recombination-mediated integration of exogenous donor DNA into a specific genomic site may also be expanded to these adult primary human cell types.
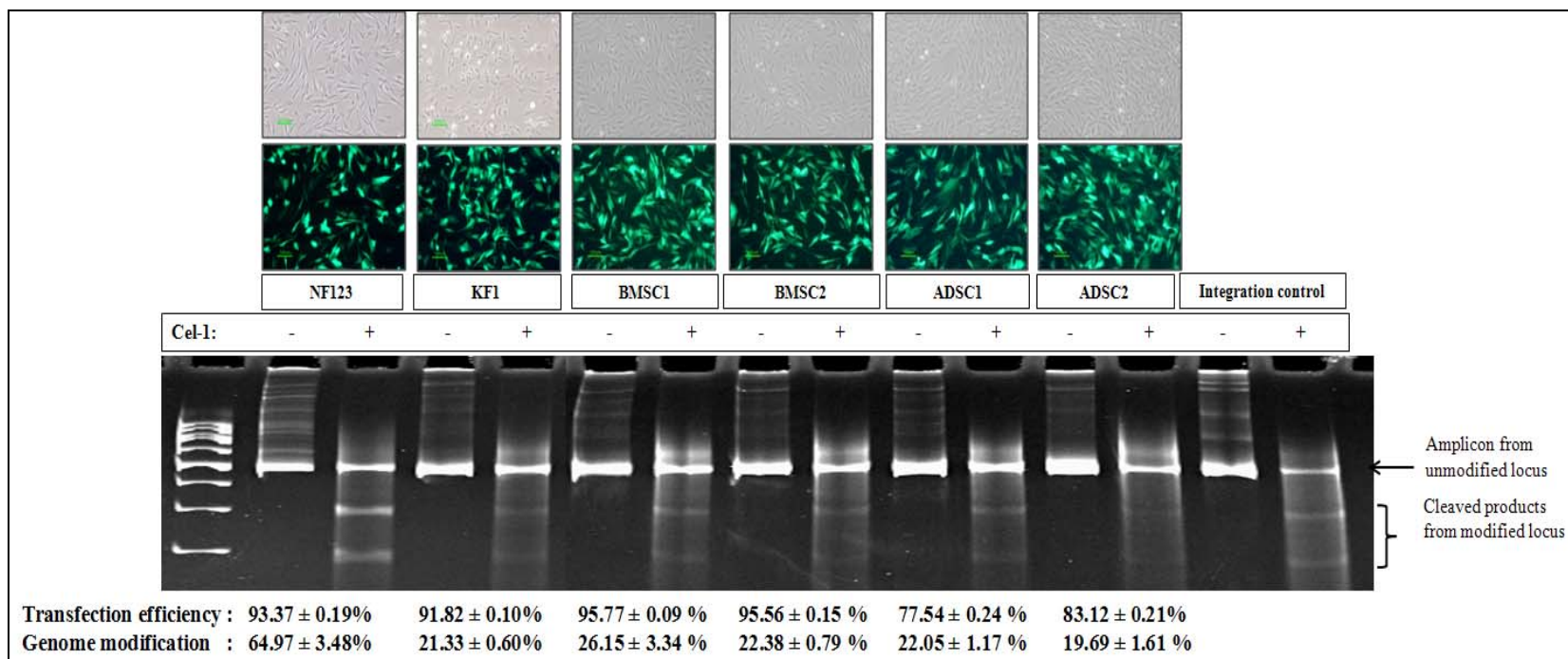
Figure 54 **Site-specific genome cleavage in primary adult human cells.** (**Top**) Brightfield and fluorescence images of primary human cells, 1 day post-electroporation with an EGFP plasmid. (**Bottom**) Genomic region spanning the AAVS1 locus was amplified from genomic DNA extracted from BMSCs (samples #1 and 2), ADSCs (samples # 1 and 2), dermal fibroblast (NF123 and KF1) and CLECs transiently electroporated with plasmid encoding for both left and right AAVS1 ZFNs (enhance Sharkey ZFNs), incubated at 37°C for 1 day followed by 30°C for 2 days. *Cel*-1 nuclease digested (+) or undigested (-) PCR amplicons were resolved electrophoretically in a 10% polyacrylamide gel, imaged and quantified using Gel Doc 2000 system and QuantityOne software. Integration control refers to PCR amplicon from a positive genomic control provided in the CompoZr® targeted integration kit (Sigma-Aldrich). Transfection efficiency determined by flow cytometry analysis of GFP-positive cells and genome modification efficiency determined by densitometry of cleaved and uncleaved amplicons are reported. Data are Mean ± SEM, n=3 per group.

263

2.4.3. Evidence of site-specific integration in other primary human cells

Given that targeted and site-specific cleavage of genomic DNA could be achieved with ZFNs specific to AAVS1 locus, we next investigated BMSC (BMSC1 and Gan BM), ADSC (ADSC1) and dermal fibroblasts (KF1) for their capacity to attain site-specific integration of a 50-bp donor DNA at the AAVS1 locus following co-electroporation with pZDonor DNA (provides 50 bp donor fragment for integration) and with (+) or without (-) plasmid DNA encoding AAVS1 enhanced Sharkey ZFNs (**Figure 55**). Junction PCR performed with a pair of vector specific and genome specific primers indicated the presence of the expected amplicon in all cell types tested when co-electroporated with DNA encoding ZFN, indicating that homologous recombination-mediated integration of an exogenously provided DNA could be achieved in these primary human cell types. Control genomic PCR with genome specific primers amplified an AAVS1 specific locus present in both modified and unmodified cells and served to demonstrate integrity of genomic DNA. Genomic DNA from K562 cells co-electroporated with the same pZDonor and ZFN plasmid DNAs served as a positive control (+ve) for integration junction PCR. It is noteworthy that the relatively weak amplicon signals in all four cell samples tested may indicate that site-specific integration is rare in these cell types under the conditions we employed. Due to the low efficiency of site-specific integration, it was not possible to confirm integration of the 50-bp DNA by RFLP assay. Further optimization will be required to increase the efficiency of site-specific integration of donor DNA in these cell types if they are to be developed for gene targeting applications.
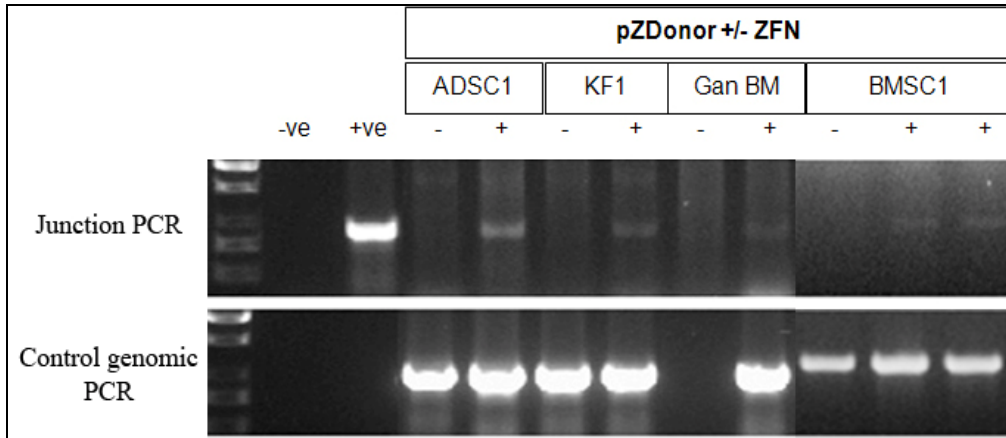
Figure 55 **Site-specific integration of donor DNA in other primary human cell types.** Two hundred ng of genomic DNA extracted 4 days post-electroporation from human adipose-derived stromal cells (ADSC1), human dermal fibroblasts (KF1) and human bone marrow-derived stromal cells (Gan BM and BMSC1) co-transfected with 10 µg of pZDonor and with (+) or without (-) 5 µg of enhanced Sharkey ZFNs was used for investigation of site-specific integration by junction PCR. "-ve" Refers to minus template amplification while "+ve" refers to amplification of a K562 genomic DNA sample previously identified to be positive for pZDonor site-specific integration. Control genomic PCR amplified a 900-bp region of the AAVS1 locus. Amplified products were electrophoresed on 1% agarose gels and imaged using Gel Doc 2000 system and QuantityOne software.

2.4.4. FVIII secretion from modified primary cells

Human BMSC, human ADSC, human CLEC and canine BMSC co-electroporated with plasmids encoding EGFP and hybrid BDD-FVIII cDNA expressed from a CMV promoter were investigated their capacity to secrete FVIII (**Figure 56**). Gene transfer efficiency (% GFP-positive cells in parenthesis) was estimated by FACS analysis to be: human BMSC1 (77.25 % ± 0.165), BMSC2 (30.08 % ± 0.187), ADSC2 (39.9 % ± 0.817), canine BMSC (73.27 % ± 0.224) and human CLECs (29.23 % ± 0.344) (Data are mean ± SEM, n=3). FVIII activity in overnight conditioned media of transfected BMSC1 (0.101 ± 0.0096 Units/ 10^6 cells/ 24 hr), BMSC2 (0.033 ± 0.005 Units/ 10^6 cells/ 24 hr), ADSC2 (0.056 ± 0.001 Units/ 10^6 cells/ 24 hr), CLEC (0.153 ± 0.0168 Units/ 10^6 cells/ 24 hr) and canine BMSC (0.179 ± 0.011 Units/ 10^6 cells/ 24 hr) measured using the Coamatic® chromogenic assay revealed that all the tested cell types were capable of processing and secreting FVIII. No FVIII activity was detected in media of untransfected wild type cells except from canine BMSC where trace levels of FVIII activity was detected (0.007 ± 0.0003 Units/ 10^6 cells/ 24 hr).

The capacity for adult primary cell types to secrete FVIII provides optimism that these cell types may potentially be developed for future gene- and cell-based therapy for adults with hemophilia A. The demonstration of FVIII secretion from canine cells opens up possibilities for testing cell-based approaches in natural canine models of hemophilia A before embarking or not on human clinical trials. Lastly, site-specific ZFNs that target a safe harbor in the genome could be used to engineer primary human cells that may prove to be safe and capable of durable secretion of transgenic FVIII.
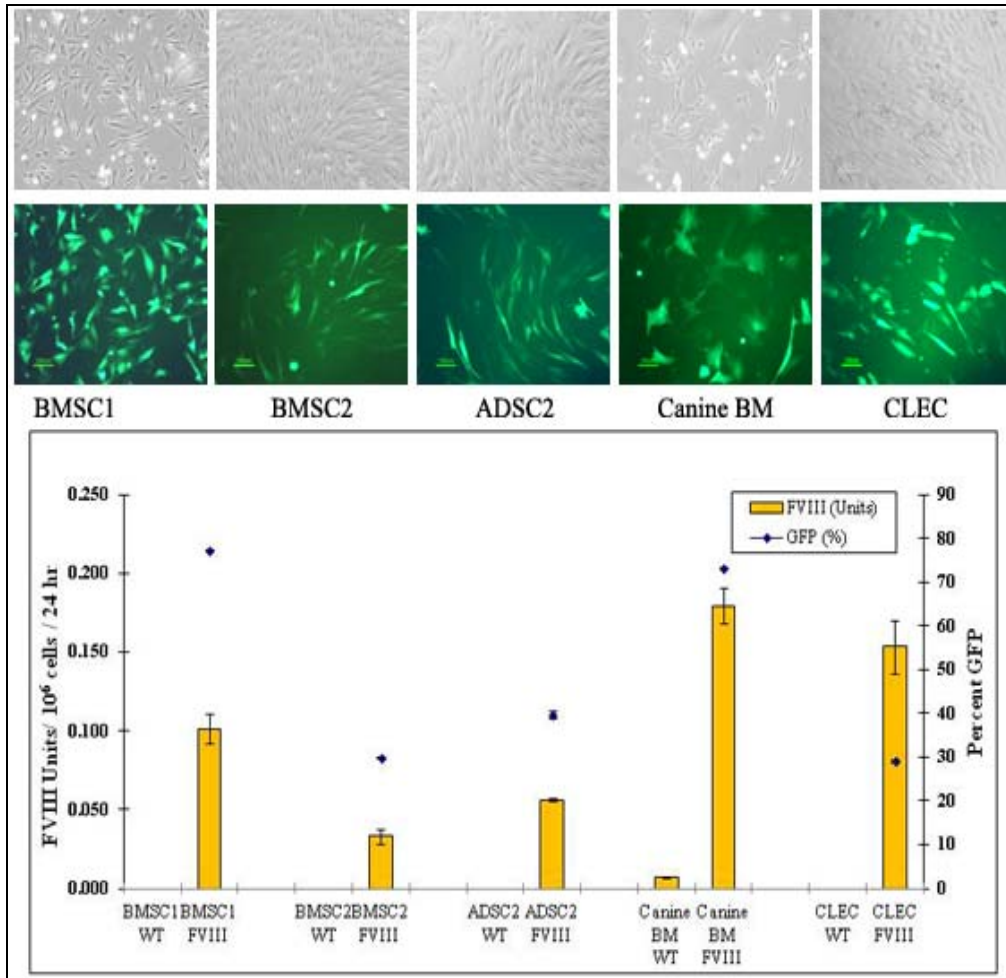
Figure 56 **Comparison of FVIII secretion in different primary cell types.**
Primary human bone marrow stromal cells (BMSC1 and BMSC2), human adipose-
derived stromal cells (ADSC2), human cord-lining epithelial cells (CLEC) and canine
bone marrow-derived stromal cells (canine BMSC) were co-electroporated with 2 μg
of an EGFP reporter plasmid and 15 μg of plasmid encoding hybrid BDD-FVIII
cDNA expressed from a CMV promoter. (**Top**) Brightfield and fluorescence images
of the electroporated cells. (**Bottom**) The percentage of GFP-positive cells (diamond
symbols) determined by FACS analysis and FVIII activity (bars) in overnight
conditioned media determined by the Coamatic® chromogenic assay and expressed as
units/ 10^6 cells/ 24 hr. Data are the means of triplicate wells.

Chapter 3

Discussion

3.1. Rationale for *ex vivo* cell and gene based-therapy for hemophilia A

Gene and cell-based therapy, a scientific concept mooted less than 50 years ago, has forayed into the frontiers of medical science as potential treatment options for genetic diseases which otherwise have no effective or affordable treatments. Evidence of clinical successes for different ailments such as a SCID-IL2RG[1], SCID-ADA[2], WAS[3], CGD[4], ALD[5], β-thalassemia[6], Leber's congenital amaurosis[7] and hemophilia B[8] have rejuvenated interest in developing gene and cell-based therapies for other genetic disorders that could equally benefit. However, iatrogenic adverse events that have surfaced even in successfully treated patients' have raised serious concerns regarding potential genotoxicity risks associated with current clinically utilized gene therapy vectors and treatment modalities[3, 9-12]. It is generally accepted that oncogenic events observed in most clinical gene therapy studies were a consequence of insertional mutagenesis and oncogenesis induced by integrating viral vectors that were used in clinical trials[13]. Given the propensity of most integrating viral vectors to preferentially integrate into transcriptionally active regions within the genome[14, 15], the risks of transactivation of oncogenes or inactivation of tumor suppressor genes remains a real concern for most integrating viral vectors. While non-integrating viral vectors such as the adenoviral and AAV vectors have shown some measure of success in clinical trials, risks of immunological complications persist with the *in vivo* applications of some of these vectors[9, 16].

With these concerns in mind, we explored an *ex vivo* cell therapy approach to modify primary human cells using non-viral means to stably integrate a transgene of interest for durable transgene expression. An *ex vivo* based strategy would allow for the comprehensive evaluation of modified cells for potential genotoxicity before they are advanced into clinical studies. The potential for immune complications associated with *in vivo* delivery of vectors would also be averted. In our study, we evaluated two different non-viral strategies to stably and precisely modify the genome of primary human cells. In the initial study, we investigated phiC31 integrase-mediated sequence-specific transgene integration into limited sites in the genome. We later progressed to ZFN-mediated site-specific transgene integration into a single targeted genomic site. Targeted transgene integration holds potential for circumventing most of the concerns associated with random integrations and the attendant risks of mutagenesis associated with most viral and even non-viral vectors. Genome-modified

cells were comprehensively evaluated for potential genotoxicity arising either as a consequence of genome modification or from long-term culture *in vitro*. In our study, we evaluated the effects of transgene integration on the global transcriptome of cells and employed high resolution techniques to investigate the genomic architecture of genome-modified cells to determine if there were amplifications or deletions of genomic regions. Genome-modified cells were spectrally karyotyped to check for gross chromosomal abnormalities. Tumorigenic potential of cells was evaluated by comparing their growth properties *in vitro* by colony formation assay and *in vivo* by their capacity to form tumors when implanted into immunodeficient mice. Collectively, in our study, we aimed to provide evidence for lack of tumorigenicity and drastic alteration of the transcriptome or genomic environment as a result of the gene-modification process.

We chose hemophilia A as a disease model for developing a gene and cell-based therapy for durable expression of FVIII from genetically modified primary human cells. Although several studies have demonstrated the feasibility and efficacy of utilizing *in vivo* viral vector delivery for FVIII expression in hemophilic animals[17, 18] and to a lesser extent in human subjects[19], issues pertaining to the use of viral vectors remain. *Ex vivo* cell and gene therapy has also been evaluated for hemophilia A. To date, several cell types have been genetically modified and tested; these include bone marrow-derived stem cells[20], blood outgrowth endothelial cells[21], long-term hematopoietic repopulating cells[22] and skin fibroblasts[23], among others. Generally, the choice of cells used for *ex vivo* cell therapy approaches depends on the disease being treated, with a preference to use gene-modified cells that are as similar as possible to those causing the disease phenotype. The rationale here would be to ensure the proper processing and expression of the transgene products from gene-modified cells. However, factors such as the availability ease of isolation and culture of these cells as well as the capacity to derive sufficient quantities for clinical applications usually broadens the choices to include other cell types as surrogates for efficiently and properly expressing the deficient transgene products.

Our study focused on a novel cell type that can be consistently and readily obtained in quantity from the lining membrane of umbilical cords. These cells, termed CLECs, were epithelioid, expressed some key markers of pluripotency and demonstrated a capacity for self-renewal and long-term propagation in culture[24, 25]. The ability to derive an estimated 6×10^9 cells from a single healthy cord even before expansion in culture indicates the feasibility of producing clinically relevant quantities of CLECs for cell therapy. These characteristics, coupled with their lack of

tumorigenicity and ethically uncontroversial derivation, make CLECs a suitable cell type for developing *ex vivo* autologous cell therapy mainly for pediatric patients with hemophilia. We envisioned that if proven successful in CLECs, the same genome modification approach could be expanded to adult primary human cell types for treating adult patients.

The inherently poor expression of human FVIII due to mRNA instability[26], transcriptional inhibitory sequences within the FVIII cDNA[27] and highly complex processing and secretory process that requires FVIII transversing from the endoplasmic reticulum to the Golgi apparatus[28, 29] have been well documented. Recombinant FVIII is known to be secreted 10^4-fold lower than other recombinant proteins of similar size[30, 31]. In an attempt to increase FVIII expression, we firstly introduced the F309S mutation[31] to human FVIII cDNA which was shown to reduce the ATP requirements for FVIII release from its association with its chaperone, immunoglobulin-binding protein (BiP), in the endoplasmic reticulum and thereby increased FVIII secretion by 3-fold[31]. A second modification was done to retain 8 N-glycosylation sites within the B-domain while deleting the rest of the B-domain of human FVIII cDNA. This modification improved FVIII secretion by ensuring its association with LMAN1 (lectin mannose binding-1) that facilitates ER-Golgi transport of FVIII protein. When comparing human and porcine FVIII cDNA, Doering *et al*., identified that porcine A1 and A3 domains conferred higher FVIII expression levels[32]. In a later phase of our study, we assembled and tested a porcine-human hybrid BDD-FVIII cDNA having 93% similarity at the amino acid level to human FVIII cDNA, and demonstrated up to 5-fold improved FVIII secretion compared to the bioengineered BDD- human FVIII F309S cDNA that was used in our earlier experiments. Although the precise molecular mechanism for improved FVIII secretion from porcine FVIII is not well elucidated, studies by Doering *et al*. point to post-translational events in the ER such as reduced unfolded protein response that may account for superior secretion compared to human FVIII[33]. Recombinant porcine FVIII has been used successfully in the past for treating human hemophilic subjects who developed high titers of inhibitory antibodies[34]. Human and porcine FVIII proteins have similar immunogenicity in animal models[35]. Thus, it is possible that the hybrid FVIII used in this study may not be more immunogenic than its human FVIII counterpart.

3.2.   PhiC31 integrase study

The use of phiC31 integrase to integrate transgenes into limited pseudo *attP* sites within the genome has been well documented[36]. In our initial study, the phiC31 integrase was utilized to integrate CMV-driven GFP reporter gene, human ferritin promoter driven BDD-human FVIII F309S and human/porcine hybrid FVIII expression cassettes into the genome of CLECs. Durable FVIII expression consistent with transgene integration was observed *in vitro* for more than 2 months in genome-modified CLECs. PhiC31 integrase has been predicted to mediate site-directed integrations into an estimated 370 pseudo-*attP* sites in the human genome, although there is evidence that this actually occurs in only a small subset of sites[37]. In this study we identified 44 independent integration events and confirmed the previously reported conserved sequence motif[37] at a majority of integration sites. Forty per cent of recovered integrations occurred at 8p22, confirming the observations made when a therapeutic transgene, *COL7A1*, was integrated into human primary epidermal progenitor cells[38]. Our data suggest that the 8p22 integration site is likely to be safe as no chromosomal translocations were detected in >132 metaphases from 5 clonal populations bearing 8p22 integrations and no changes were detected in the expression of genes that mapped within 1 Mb intervals centered around 8p22 integrations. This tendency to integrate at 8p22 could be further enhanced by developing integrase variants of higher target site specificity[39, 40]. Given the high propensity to integrate at 8p22, we embarked on a study to further evaluate genotoxicity potential of clonal CLECs with transgene integration at 8p22. We envisioned that well characterized clonal CLECs with known integration sites would be more readily accepted for clinical applications compared to heterogeneous populations of genome-modified cells that harbor integrations at multiple genomic sites. However, given the difficulty of establishing pure monoclonal CLEC cultures by flow sorting single cells (probably due to poor survival and growth after flow sorting), our clonal studies were based on oligoclonal cells derived from initially sorting 4 cells.

One of the possible adverse effects of integrating vectors is dysregulation of the function of genes at or close to integration sites[41-43]. We determined the influence of transgene integration on the transcriptome of a bulk population of phiC31 integrase-modified CLECs. The technique employed had a detection sensitivity[44] of 1 transcript per 200000 and thus could accurately reflect transcriptional changes in even a small proportion of genome-modified CLECs. Our analysis of transcriptome data with reference to retrieved integration events revealed no significant perturbations in the expression of genes close to integration sites or within 1 Mb intervals centered around these sites, with the single exception of *LNX1*, thus confirming the

transcriptionally benign effects of these integrations. It was noteworthy that even potential oncogenes and tumor suppressor genes in these regions showed no perturbation of expression. Transcriptome analysis of 8 oligoclonal CLECs with transgene integration at 8p22 locus also did not reveal any significant changes in the expression of the endogenous *DLC1* gene or neighboring genes that mapped within 1 Mb intervals centered around the 8p22 integration site. These data are reassuring in that transgene integration at 8p22 did not alter transcriptional landscape of neighboring genes.

Global transcriptome analysis did reveal that 1.3 % of total expressed genes had significant changes ($\geq$ 2-fold) in the bulk population of genome-modified CLECs. A major biosafety concern of genome modification is the risk of mutating or dysregulating oncogenes or tumor suppressor genes[41, 45]. In our phiC31 integrase work, 15 of 151 transcriptionally altered genes were either potential oncogenes or tumor suppressor genes. Pathway mapping of these 151 transcriptionally altered genes identified only three genes that mapped significantly to a single pathway (p53 signaling; *P* = 0.02). Similarly, analysis of 8 oligoclonal CLECs with transgene integration at 8p22 locus identified 9 potential oncogenes with up-regulated expression and 31 potential oncogenes with down-regulated expression. Given that 45% of the 250 genes that had significant down-regulated expression belonged to the category of cell-cycle genes, reasonable assumptions would suggest retardation of proliferation in genome-modified CLECs as opposed to uncontrolled proliferation associated with transformed cells. The likelihood that genome-modified CLECs were transformed was judged to be low based on unaltered colony-forming activity *in vitro* and absence of tumor development in immunocompromised mice *in vivo*.

Previous reports have shown the association of phage integrase-mediated integrations with deletions of up to a few thousand basepairs and insertions of up to a few hundred basepairs at integration sites[46]. In our study however, DNA sequence analysis revealed microdeletions of vector DNA ($\leq$37bp) at recovered integration junctions. Using high-resolution genome copy number data (Affymetrix Human Mapping 500K array set) of a mass culture of genome-modified CLECs showed significant copy number changes (two deletions and one gain) in only 3 genomic regions that were not integration sites. Similarly, our evaluation of oligoclonal CLECs using a higher density SNP array (CytoScan HD array) with an average probe spacing of 1148 bp revealed no significant alteration of the genomic architecture in 7 of 8 oligoclonal CLECs that were investigated. Minor amplification of a 4118 kb centromeric region was noted in chromosome 19 of a single clone but without any

transcriptional effects on the gene in close proximity, *ZNF254*. Thus, no major deletions or insertions could be ascribed to phiC31 integrase-mediated transgene integrations and the few copy number changes detected had minimal transcriptional effects of uncertain functional relevance.

Another major concern from studies of phiC31 integrase in human cells to date is the occurrence of chromosomal translocations in up to 15% of stably modified cell lines[46], primary human embryonic and adult fibroblasts[47, 48]. However, our analysis of >300 spectral karyotypes of phage integrase modified cells revealed only 2 cells with non-recurring chromosomal translocations and 2 cells harboring the same translocation. These rearranged chromosomes were only detected in mass cultures of CLECs and none was found in 8 oligoclonal populations. The frequency of observed translocations in this study (4 of 300 metaphases) was lower compared to other reports (15% - 30%) and may be related, in part, to the different cell types used by other investigators for phiC31 integrase modification[46-48]. Surveys of two large series of prenatal genetic screening have shown *de novo* chromosomal translocations in amniocytes, chorionic villus and fetal blood samples. In one study, normal infants were born of pregnancies that were not terminated on the basis of abnormal karyotypes[49]; while in the other study, the risk of congenital malformations was sufficiently close to the background rate that it did not support a relationship of chromosomal rearrangements to somatic abnormalities[50]. While both studies are limited with respect to accurate risk assessment, it remains true that at least some *de novo* chromosomal translocations are functionally silent and inconsequential[51]. Our data are consistent with the proposition that translocations in phiC31 integrase modified cells are uncommon stochastic events that do not confer either a survival or proliferative advantage to the affected cell(s). Had this not been the case, the translocations we identified in modified CLECs would have been the dominant karyotype in a mass culture, rather than the highly sporadic events actually observed. Oligoclonal CLECs were evaluated for chromosomal aberrations using the high resolution Cytoscan HD array which is more sensitive than conventional spectral karyotype techniques which have a resolution[52] of around 1Mb. Our array data revealed no evidence of gross deletions, insertions or translocations in all 8 oligoclonal CLECs that were analyzed.

Although translocations are a cause for concern, it is axiomatic that malignant transformation results from multiple, rather than single, genetic and genomic alterations[53]. Moreover, specific rather than random translocations are associated with hematologic and solid tissue malignancies [54]. Robust defenses against neoplastic

transformation in the form of cell cycle arrest, apoptosis and senescence are activated by the genotoxic stress of unrestrained cell proliferation induced by mutations and/or genomic aberrations[55]. Operation of these innate tumor suppressive mechanisms probably explain the fact that mass cultures of phiC31 modified CLECs did not develop clonal dominance of cells harboring chromosomal translocations.

The benign safety profile of phiC31 integrase modified CLECs suggested by *in vitro* analyses was confirmed by lack of *in vivo* tumorigenicity when these cells were implanted into immunocompromised NOD-SCID mice. Oligoclonal CLECs with transgene integration at 8p22 locus also showed no evidence of tumorigenicity for at least 4 months even when implanted into the more severely immunodeficient NSG strain of mice, whereas parallel implantation of a tumorigenic epithelial cell line (Hs746T) resulted in tumor formation within 3 weeks. Although there was evidence of loss of implanted cells as indicated by declining levels of transgenic FVIII, immunohistochemical staining for human vimentin did show the presence of engrafted CLECs in explants as well as from cultures re-established from explants. Given that tumor formation could be observed in NOD-SCID mice implanted with as few as 100-1000 cancer cells[56, 57], tumorigenic CLECs even if present in small numbers should have given rise to tumor formation in NSG mice. Failure of implanted CLECs to form tumors reiterates other data pointing to their intrinsic lack of tumorigenicity.

Implantation of human FVIII-secreting CLECs raised plasma FVIII levels and partially corrected the bleeding phenotype in hemophilic mice, suggesting the potential for developing non-genotoxic cellular therapy that could be especially effective for autologous or allogeneic applications[58]. Modest increase in plasma FVIII levels could be attributed to the known short circulating half-life of human FVIII in mice (74 minutes cf. 12 hours in humans)[59] and sub-optimal engraftment and vascularization of implanted human cells in a xenogeneic host. Oligoclonal CLECs secreting hybrid FVIII resulted in 10-fold higher plasma FVIII levels even when implanted with 2-fold fewer cells into NSG mice. This is consistent with higher FVIII levels attained with hybrid FVIII compared with human FVIII and possibly lower immunogenicity in NSG mice. The eventual decline of plasma FVIII levels to background levels by day 30 could have been due to loss of implanted cells because of sub-optimal engraftment and vascularization in a xenogeneic host. A comprehensive evaluation of durable FVIII expression *in vivo* will require long-term engraftment of genome-modified cells in an autologous model such as testing autologous canine cells in a canine model of hemophilia.

Despite being successfully utilized for genome modification of several mammalian cell types, the phiC31 integrase has not made much progress towards clinical applications. The potential to induce chromosomal translocations and the capacity to mediate transgene integrations into multiple genomic sites are major limiting factors associated with the use of phiC31 integrase. In this study, we show that phage integrase modified CLECs expanded as polyclonal mass cultures appear to have sustained few or no potentially oncogenic genomic alterations. This approach could be rendered even safer by implanting clonal populations such as those with integrations only at 8p22 locus, pre-screened *ex vivo* for biosafety using a range of assays such as those we employed. Moreover, our demonstration that phiC31 integrase can modify primary adult human cells such as fibroblasts, ADSCs and BMSCs to stably secrete FVIII *in vitro* opens up possibilities of expanding this approach for developing gene and cell -based therapy for adult hemophilic patients. The relatively high frequency and safety of transgene integrations into the 8p22 locus also opens up the possibility for screening and selecting clonal cells with this particular site-specific integration. The use of integrases with greater specificity[39, 40], high-throughput screening methods for selecting safe clones[60] and the use of cells with high proliferation capacity could make this approach more acceptable for clinical adoption.

3.3. ZFN study

Currently utilized integrating gene therapy vectors mostly integrate randomly and into multiple sites within the genome, thus incurring significant risks of inducing mutagenesis and oncogenesis. Vector systems such as Sleeping Beauty transposons and phiC31 integrase should only be considered quasi-site-specific as integrations have been observed to occur into sites with little conserved sequence specificity. Our interest in zinc finger nucleases arose from their supposed capacity to target and cleave unique sites within the genome and to mediate site-specific integration of exogenous donor DNA by homologous recombination.

The choice of genomic region for targeted transgene integration is an important consideration given the propensity of transgene integration to activate or inactivate endogenous and/or neighboring gene(s) at or in the vicinity of sites of integration. The AAVS1 locus has been proposed as a potential safe harbor within the human genome given the lack of pathological consequences in patients with natural AAV2 viral vector integration at this locus[61]. No functional consequences have been reported from inactivation of the endogenous *PPP1R12C* gene[62, 63], whose function

has not been fully elucidated but is thought to involve the regulation of actin-myosin fiber assembly as the protein appears to be a regulatory subunit of a myosin-binding protein phosphatase[64]. Furthermore, it has been suggested that natural insulator elements at AAVS1 could prevent transgene silencing and transactivation of neighboring genes[65]. Several studies have supported this notion by demonstrating durable transgene expression from genomic integration of transgenes at the AAVS1 locus[62, 63, 66]. We investigated ZFNs specific for the AAVS1 locus to site-specifically integrate a human ferritin promoter driven human-porcine hybrid FVIII cassette into the AAVS1 site of CLECs with the intent of developing these cells as durable FVIII secreting bioimplants for gene and cell-based hemophilia treatment. This part of the study primarily focused on evaluating the efficiency of targeted transgene integrations at the AAVS1 locus, durable transgene expression from targeted transgene integrations, the extent of off-target cleavage by ZFNs and potential and actual genotoxicity associated with transgene integration.

Successful genome modification with the use of ZFNs depend on several factors such as the quality and specificity of the ZFNs used, high gene transfer efficiency, and accessibility of the ZFN protein to the genomic region of interest The latter, in turn, is influenced by chromatin configuration of the genomic region (condensed or heterochromatin configuration *versus* open euchromatin configuration), intrinsic capacity of the cells to repair DNA damage by either NHEJ or HR pathways and tolerance to cellular toxicity from expressed ZFNs.

ZFNs have been designed and constructed using various strategies such as the modular assembly method[67], oligomerized pool engineering (OPEN) assembly method[68] and context-dependent assembly (CoDA) method[69]. Commercial service providers are another source of customized ZFNs, such as Sangamo Biosciences whose proprietary design schemes (CompoZr) may be contracted to assemble highly specific ZFNs. Given that the success rate of designing a highly efficient and specific ZFN varies with the different methods of assembly, we chose to assemble our ZFNs based on a previous publication from Sangamo Biosciences which showed high targeted genome cleavage activity[63]. The AAVS1 ZFN pairs assembled in our lab were codon-optimized and expressed as a single plasmid vector to improve the stoichiometric expression of each monomer in mammalian cells. We further modified the catalytic subunits of *Fok*I endonuclease to have higher nuclease activity[70, 71]. These modifications to ZFNs in combination with transient mild hypothermic incubations[72] indeed resulted in more efficient genomic cleavage activity probably by retarding ZFN protein degradation in K562 cells, CLECs and other primary human

cell types. In cell types that require higher levels of ZFN activity, it may be possible to further increase ZFN activity by expression from cell type specific cellular promoters such as the human ferritin promoter used for expression of FVIII expression in CLECs. The use of proteasome inhibitors such as MG132 to reduce ZFN degradation and thus increase ZFN activity has also been reported as another feasible strategy[73].

Efficient expression of ZFNs is predicated on achieving high gene transfer efficiencies in the cell types tested. Given the inherent difficulty of transfecting primary human cells, each cell type used in this study had to be first optimized for gene transfer efficiency using an EGFP reporter gene and the Amaxa®Nucleofector™ system. Gene transfer efficiencies of greater than 60% were achieved in most cells types used in this project. While this should ensure higher levels of ZFN expression and activity, intriguingly, in some cell types such as fibroblasts, ZFN activity did not correlate with gene transfer efficiencies. Thus, factors intrinsic to cell types and sources of cell types also influence ZFN activity.

It is plausible to anticipate that ZFN binding and activity in different cell types may be influenced partly by epigenetic factors such as the chromatin configuration of the genomic site being targeted. Chromatin accessibility may be a critical determinant of ZFN binding and activity, and may explain the differences in genome modification activities reported in different cell types. Thus, there could be a higher likelihood of targeting ZFNs to a euchromatic locus with an active chromatin configuration compared to heterochromatic regions with predominantly inactive chromatin configuration[74, 75]. However, it is noteworthy that even transcriptionally silent loci have been modified by ZFNs[63].

One of the current limitations of using ZFNs to mediate targeted transgene integration is the relatively low frequency of HR achieved in primary human cell types. Differences in the intrinsic capacity for NHEJ or HR repair could explain differences in the frequency of gene insertion or indel formation in different cell types treated with ZFNs. The phase of the cell cycle during which DNA damage is detected and repair initiated has been thought to influence the choice of repair pathways as well. Generally, the NHEJ pathway which may result in indel formation predominates in G1 although it is known to function throughout the cell cycle, whereas HR is thought to occur exclusively during late S to G2/M phases[76, 77]. Transient cell cycle arrest and cell cycle synchronization have been experimentally shown to alter the ratio of HR/NHEJ mediated repair following double-stranded DNA breaks[76, 77] and may be manipulated to enhance HR mediated transgene integration in ZFN studies[78].

ZFNs modified to function as single-stranded DNA nicking enzymes have been shown to favor HR over NHEJ repair, and could yet be another strategy for enhancing HR frequencies for targeted gene integration studies[79, 80].

Cellular toxicity arising from ZFN expression in cells has been reported in some studies. While part of the toxicity has been ascribed to the off-target effects on the genome presumably when functioning as homodimers, cellular toxicity may also be related to the intrinsic capacity of cells to tolerate and repair double-stranded breaks, and could explain differences in toxicity observed in different cell types. In our study, ZFNs specific to the AAVS1 locus[63] were cloned and optimized to function as obligate heterodimers[81] to reduce off-target genomic effects and mitigate cellular toxicity. Transfection studies using obligate heterodimer ZFNs showed that CLECs could tolerate ZFNs at doses of 5 to 10 μg without significant cellular toxicity as determined by monitoring EGFP-expressing cells, quantification of phosphorylated histone H2AX (a marker for double–stranded DNA damage) and by MTS assay. Modulating ZFN levels with small-molecule proteasome inhibitors and further improvements to heterodimer ZFN variants have also been proposed as ways to further minimize ZFN toxicity[82]. Unlike the phiC31 integrase study, no gross chromosomal abnormalities were observed with the use of ZFNs. Analysis of 23 metaphases by spectral karyotyping revealed no chromosomal abnormalities in stable CLECs with transgene integration at the AAVS1 locus.

The capacity for AAVS1 ZFNs to target, cleave and repair the AAVS1 genomic locus by either NHEJ (in the absence of donor DNA) or HR (in the presence of exogenous donor DNA) were evaluated by *Cel*-1 mismatch nuclease assay and RFLP assay, respectively. *Cel*-1 nuclease assay has been estimated to have detection sensitivity[83] of 1:32. Having determined the capacity to target and cleave the AAVS1 genomic region and induce homologous recombination of short donor DNA (50-bp), we were able to demonstrate site-specific integration of donor DNA of progressively increasing size i.e. 1-kb, 4-kb and 9-kb. Positive selection (either neomycin or puromycin) was usually necessary to enrich for cells with integration of donor DNA. Site-specific donor DNA integrations were reliably detected by integration junction PCR using primer pairs anchored to the genomic region and integrated vector and confirmed by sequencing. ZFN-mediated HR repair of donor DNA was found to be highly accurate with no evidence of deletion of either vector or donor DNA at the integration junctions. In addition to integration junction PCR, we employed long-range PCR and sequencing to evaluate the complete integration of 1-kb, 4-kb and 9-kb donor DNA. Sequencing of these genome amplified regions showed no evidence

of deletion or rearrangements within the vector itself, reiterating the accuracy of HR mediated repair. However, in our ZFN study, we did encounter targeted transgene integration events in which only one integration junction could be detected, raising the possibility of partial or incomplete integration. It should be emphasized that these anomalous targeted integration events were only observed in experiments where the neomycin resistance gene was expressed from an exogenous promoter. However, both integration junctions were readily detected in later experiments that used the promoter trap strategy to express the puromycin resistance gene from the endogenous *PPP1R12C* promoter. We postulate that expression of the positive selection gene from an endogenous cellular promoter was a more stringent technique for selecting cells that had integrated the transgene at the correct locus. In contrast, expression of the positive selection gene from its own exogenous promoter selected cells with random integration as well as targeted integration events, and even when incomplete integration may have occurred. Thus, the use of a promoterless puromycin selection construct was effective in reducing mis-selection of cells with undesired genotypes. We used an end-point PCR method and densitometric measurements to quantitate and estimate the ratio of on-target integration and total integration. These methods generally suffer from low sensitivity and lack of dynamic range of detection and thus should be supplemented with more sensitive and specific methods such as real-time PCR or digtal PCR.

The use of suicide gene expression for negative selection could also serve as a complementary strategy to eliminate cells with random integration. Cells with targeted integration would not be expected to incorporate the suicide gene whereas cells with random transgene integration would be expected to express the suicide gene whose expression is used to convert a non-toxic prodrug into a toxic cytotoxic agent. We explored the use of a codon-optimized HSV-TK007 suicide gene[84] to eliminate cells with random integration when combined with gancyclovir selection. CLECs expressing codon-optimized HSV-TK007 were more effectively eliminated by gancyclovir compared to cells expressing HSV-TK. However, we noted that when positively selected (with neomycin) CLECs with transgene integration were exposed to gancyclovir, surviving cells entered into a non-dividing senescence-like state. Further evaluation and modification of such a positive-negative dual selection strategy will be required in CLECs, if considered necessary. An alternative suicide system such as one utilizing the caspase system to induce apoptosis[85] could be more effective in eliminating undesirable cells. Ideally, clonal cells should be screened by FISH, genomic PCR or Southern blotting techniques to identify those which only

have targeted transgene integration and no random integrations. We have shown that CLECs could potentially be expanded from oligoclonal cell cultures and sufficient cells could potentially be derived from screening and identifying a number of safe clones. The unlimited proliferative capacity of iPS cells could be very useful in such situations where safe clonal cells are chosen for expansion provided that the risks of terotoma formation from undifferentiated iPS cells are overcome.

Off-target genome modification is another concern associated with the use of ZFNs[86]. Cleavage and repair by error prone NHEJ elsewhere in the genome besides the targeted region could result in frameshift mutations, which could be of serious consequence if it occurs within crucial regulatory or transcript-coding regions. The use of well designed and obligate heterodimeric ZFNs has been shown to significantly reduce the potential for off-target binding and activity of ZFNs[81, 82, 87]. Gaj et al.[88], described fewer off-target effects of ZFNs when delivered as purified proteins compared to expression from nucleic acids. Potential off-target effects can be bioinformatically predicted by programs such ZFN-Site[89], experimentally determined by assays such as SELEX[63] or by retrieving integration events following co-treatment with ZFNs and an episomal vector such as the integrase-defective lentiviral vector[90]. In our study of AAVS1 ZFNs, we evaluated previously identified top-10 potential off-target sites[63] for indels by deep sequencing. Our data confirm the potential for AAVS1 ZFNs to induce low frequency indels at one of the predicted off-target sites (OT1) in CLECs. Given that OT1 resides within an intergenic region distant from protein coding regions, it is unlikely that these low frequency deletions and/or insertions of very short nucleotides (<5 nucleotides) would be consequential. Previous studies utilizing ZFNs targeting the AAVS1 locus evaluated these off-target sites by the less sensitive Cel-1 mismatch assay in a different cell type and found no evidence of off-target effects. Our control studies performed by spiking a wild type amplicon with known concentrations of a synthetic amplicon with mutations (5 bp deletion) established that deep sequencing could detect mutants present in as low as 0.1% of a normal population. Thus deep sequencing could prove to be a more reliable and sensitive method for detecting rare cells with indels in a mixed population of normal cells following ZFN treatment. Evaluation of clonal cells would not be anticipated to require as high a depth of sequencing coverage for indel detection. In our study, we only evaluated the previously reported top-10 potential off-target sites. The possibility of off-target effects elsewhere in the genome cannot be excluded and whole genome sequencing would provide a complementary, comprehensive and thorough evaluation.

Besides off-target evaluation at the genomic level, we performed transcriptome analysis to determine if transgene integration had perturbed the expression of critical genes and pathways. The use of targeting vector with splice-acceptor sequence and 2A peptide down-regulated expression of the endogenous *PPP1R12C* gene as expected, and was confirmed by RT-PCR and transcriptome analysis. Our RT-PCR results showed a 50% reduction in *PPP1R12C* gene expression in CLECs with stable transgene integration compared to unmodified CLECs, suggesting that most cells were monoallelic for the integrated transgene. Although endogenous *PPP1R12C* gene expression was reduced, expression of neighboring genes within 1 Mb of the integration site remained unaffected as were the majority of potential interacting partners and downstream effector genes associated with *PPP1R12C*. Furthermore, proliferation and cellular morphology of genome-modified cells appeared normal. Studies have shown normal cell physiology and growth even with biallelic disruption of the *PPP1R12C* gene, suggesting that this gene function may not be crucial for cell survival and function or that other related genes are able to compensate for its reduced expression[91]. Others have reported that targeted transgene integration could be achieved at the AAVS1 locus without disruption of *PPP1R12C* gene expression by integrating into the opposite non-coding strand of *PPP1R12C* gene; this may be considered if it proves important to avoid disruption of the endogenous gene expression[92].

In summary, we were able to derive stable CLECs with site-specific transgene integration at the AAVS1 locus and demonstrate durable FVIII transgene expression. These genome-modified cells had normal growth characteristics, unaltered transcriptome of neighboring genes, potential interacting partners and downstream effector genes, normal chromosomal karyotype and lacked significant off-target effects at potential off-target ZFN binding sites. Our data are consistent with the notion that ZFN modified cells may be safe for gene and cell therapy applications. Further studies are required to comprehensively profile the biosafety of ZFN edited cells and will be discussed in section 3.6.

3.4.   Clinical relevance of ZFN-modified cells

The use of non-viral vectors for development of *ex vivo* cell therapy for hemophilia A treatment is a strategy that has been explored by several other groups before. Successful therapeutic effects, albeit at low levels and for up to 10 months, were reported in at least 1 clinical trial for severe hemophilia A patients that implanted FVIII plasmid transfected autologous primary fibroblasts[23]. This provided

proof-of-concept for *ex vivo* gene and cell-based therapy of hemophilia A. We postulate that the lack of durable therapeutic effect in the aforementioned clinical trial may have been due partly to random integration of the FVIII transgene cassette into unfavorable genomic loci that failed to sustain expression. Alternatively, the integrated transgene may have been silenced by prokaryotic elements present in the vector. With respect to the former, our strategy of utilizing zinc finger nucleases to site-specifically integrate the FVIII transgene into a genomic locus capable of supporting durable expression might be considered an improvement in terms of achieving biosafety as well as therapeutic durability. Further improvements to FVIII expression with the use of hybrid FVIII cDNA driven from a strong cellular promoter resulted in high expression of FVIII and may translate eventually to requiring fewer cells for implantation. Calculations based on our best FVIII expression from CLECs (2131 mUnits FVIII/ 10^6 cells/ 24 hr) estimate that approximately 100 to 400 million cells would be required to achieve a steady-state FVIII level of 10% in a 20 - 40 kg child. Given the high proliferative capacity of CLECs and the potential to derive up to 6×10^9 cells from a single umbilical cord, genome-modified CLECs could be expanded to sufficient numbers and potentially serve as useful source of autologous cells for treating pediatric and young hemophilic children. It should be noted that this could only be feasible in clinical situations where the family history alerts the parents and medical attendants to the probability of the birth of a baby with hemophilia, in which case the umbilical cord can be collected at birth and processed appropriately.

The gene targeting strategies highlighted in this project utilizing CLECs may be expanded to other adult cell types to treat adult hemophilic patients. Besides hemophilia A, other genetic ailments such as hemophilia B, epidermolysis bullosa, lysosome storage diseases such as Pompe disease and Gaucher disease and diseases caused by deficiency of a secreted protein may potentially benefit from *ex vivo* gene and cell therapy. ZFNs appear to be useful tools for mediating targeted integration of relevant transgenes for safe and durable expression in suitable adult somatic cells or stem cells for developing *ex vivo* gene and cell therapy for some of these disorders.

3.5. Conclusions

Our results showed that phiC31 integrase could efficiently integrate transgenes in a sequence-specific manner into limited sites within the human genome to generate genome-modified CLECs capable of durable FVIII expression. Genome-modified CLECs were non-tumorigenic and not associated with any overt signs of transcriptional alterations or copy number changes characteristic of transformed cells. We were able to further show that clonal CLECs with transgene integration at

chromosome 8p22 could be isolated and expanded in culture, and had few features of potential genotoxicity as ascertained by high resolution copy number change analysis, transcriptome analysis and tumorigenicity studies in immunodeficient mice.

Our study showed that CLECs were capable of secreting high levels of bioactive transgenic FVIII and demonstrated potential for expansion to very large cell numbers, even from very modest initial cell numbers, reiterating their potential value for cell therapy applications. Consistent with previous studies, we were able to confirm that transfecting a hybrid FVIII cDNA consisting of porcine and human FVIII domains resulted in cells secreting 5- to 10-fold higher levels of FVIII compared to human FVIII cDNA.

In the second part of our study, we successfully demonstrated ZFN-mediated site-specific integration of transgene cassettes in CLECs. AAVS1 ZFNs were highly site-specific and lacked significant off-target genome cleavage activity of concern. Integration of hybrid FVIII transgene cassette into the AAVS1 locus of CLECs resulted in durable FVIII secretion without affecting the expression of neighboring genes, and with very little or no perturbation of potential interacting partners and downstream effector genes of *PPP1R12C*. ZFNs may be useful agents for site-directed transgenesis in a broader range of primary human cell types such as fibroblasts, bone marrow- and adipose-derived stromal cells that could be developed for *ex vivo* cell therapy.

In conclusion, CLECs are a useful cell type for developing *ex vivo* cell therapy applications for pediatric patients. Hybrid FVIII cDNA resulted in superior FVIII production compared to human FVIII cDNA. Stable FVIII expression could be achieved by genomic integration of FVIII transgene cassette using either phiC31 integrase or ZFNs. Whilst both systems showed low genotoxicity potential, the ZFN system which is highly site-specific and allows user-defined selection of genomic integration site could be superior. ZFNs merit further development as useful tools for developing non-viral *ex vivo* cell therapies.

3.6. Future work

In this study, we demonstrated that CLECs could be modified using the phiC31 integrase and ZFN systems to stable integrate and durably express FVIII transgene. We further show that the same genome modifying agents could be applied to other primary human cell types such as fibroblasts, bone marrow- and adipose-derived stromal cells, which are more appropriate for developing cell-based therapy for adult patients. Most of the work in this project was directed at modifying CLECs, a cell type mainly relevant for treating pediatric patients. Therefore one direction to pursue

in future is to extend these approaches to primary cell types that are readily procured from human adults, as mentioned above. As ZFNs are superior to phiC31 integrase in their ability to modify unique user-specified sites in the genome, it would make more sense to focus on the capacity of ZFNs to modify primary adult human cell types. The choice of relevant cell types must not only be made on the basis of the ability to efficiently modify their genomes but also on the capacity to expand genome-modified cells to levels that are clinically relevant. In this respect, it would be of interest to include iPS cells in future work, as these cells have unlimited proliferation potential under appropriate conditions of culture. iPS cells can be derived and used in an autologous manner and would not be limited by factors such as the age of the donor and cell senescence. Furthermore there are now efficient methods to generate iPS cells using techniques that do not directly modify the genome. In theory, clonal iPS cells having been comprehensively evaluated for low or no genotoxicity may be expanded to derive sufficient cell numbers and differentiated into a relevant cell type for therapeutic applications, although risks pertaining to teratoma formation from undifferentiated iPS cells must be overcome. Genotoxicity studies on ZFN modified cells would be more meaningful if complemented with whole genome sequencing to screen for off-target effects of ZFNs on an unbiased genome-wide scale.

Aside from potential off-target effects, another issue associated with the use of ZFNs is the generally low efficiency of double-stranded break achieved at targeted genomic sites in primary human cell types. Thus, it would be useful to compare the genome editing effects of other site-directed nucleases such as the TALE nucleases[93] and RNA-guided CRISPR-Caspase9 systems[94]. Besides comparing the genomic cleavage activity of these nucleases, it will be important and informative to compare potential off-target effects among the different genome-editing methods. The capacity to efficiently and accurately create site-directed DNA breaks in the genome could be combined with HR promoting strategies, such as cell-cycle synchronization and creation of single-stranded DNA nicks instead of double-stranded DNA breaks to enhance targeted gene integration.

Current animal studies based on xenogeneic implantation of gene-modified human cells into immunodeficient mice are not ideal for long-term tumorigenicity assessment due to failure to achieve long-term engraftment of sufficient number of cells either due to immune response or lack of vascularization. Long-term monitoring of implanted cells would thus be ideal in an autologous setting where the issues related to immune response are absent. As such an ideal model to show efficacy and safety for *ex vivo* cell therapy for hemophilia A would be one based on autologous

implantation of gene modified canine cells into hemophilic dogs. Bone marrow cells, fibroblasts and iPS cells could be suitable cell types for these studies. However, gene modifying agents targeting a suitable locus in canine cells need to developed *de novo* and tested in order to proceed with these studies. Studies in a large animal model will also test the feasibility of scaling up autologous cell therapy for human applications.

A final outstanding issue with genome modification is the identification and testing of potential safe genomic harbors for targeted gene integration in addition to the AAVS1 site. Some studies have noted transcriptional dysregulation to either endogenous genes at the sites of integration or to neighboring genes close to integration sites[91, 92, 95], although no untoward effects have been reported. It would be useful to investigate other genomic loci, such as the 8p22 region (intron 1 of *DLC1*) that we had highlighted in our integrase study, where transgenesis and durable transgene expression can be achieved without effects on endogenous or neighboring genes. Whole transcriptome analysis should be performed using more up-to-date technologies such as RNASeq[96] to yield more comprehensive and accurate reflection of transcriptional changes in genome modified cells. The inclusion of insulator elements within the integrating vector may be necessary to prevent transgene silencing as well as effects on neighboring genes. The overall strategy of integrating transgenes into the genome could benefit from inclusion of reliable suicide genes[97] into the integrating vector, to be effectively activated when and if necessary to eliminate rogue cells.

In summary, future work should focus on testing different adult human cell types for amenability to genome modification, should evaluate different genome modifying agents and different genomic loci for safe and efficient transgenesis. Whole genome sequencing may be crucial for evaluating off-targets of nucleases as an unbiased genome-wide assessment. Long-term genotoxicity monitoring and evaluation of *in vivo* efficacy should ideally be tested in an autologous manner in large pre-clinical animals such as the canine models before transiting to human clinical trials.

3.7. References

1. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F*, et al*. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science. (2000) 288:669-672.

2. Aiuti A, Cattaneo F, Galimberti S, Benninghoff U*, et al*. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med. (2009) 360:447-458.

3. Boztug K, Schmidt M, Schwarzer A, Banerjee PP*, et al*. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. N Engl J Med. (2010) 363:1918-1927.

4. Ott MG, Schmidt M, Schwarzwaelder K, Stein S*, et al*. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med. (2006) 12:401-409.

5. Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Bougnères P*, et al*. Lentiviral hematopoietic cell gene therapy for X-linked adrenoleukodystrophy. Methods Enzymol. (2012) 507:187-198.

6. Cavazzana-Calvo M, Payen E, Negre O, Wang G*, et al*. Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. Nature. (2010) 467:318-322.

7. Simonelli F, Maguire AM, Testa F, Pierce EA*, et al*. Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. Mol Ther. (2010) 18:643-650.

8. Nathwani AC, Tuddenham EGD, Rangarajan S, Rosales C*, et al*. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med. (2011) 365:2357-2365.

9. Raper SE, Chirmule N, Lee FS, Wivel NA*, et al*. Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. Mol Genet Metab. (2003) 80:148-158.

10. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, *et al*. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest. (2008) 118:3132-3142.

11. Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, *et al*. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest. (2008) 118:3143-3150.

12. Stein S, Ott MG, Schultze-Strasser S, Jauch A, *et al*. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med. (2010) 16:198-204.

13. Baum C. Current opinions in hematology 2007 vol 14 Insertional mutagenesis in gene therapy. Hematology. (2007) 14:337-342.

14. Modlich U, Navarro S, Zychlinski D, Maetzig T, *et al*. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther. (2009) 17:1919-1928.

15. Mitchell RS, Beitzel BF, Schroder ARW, Shinn P, *et al*. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. (2004) 2:E234.

16. Chuah MKL, Collen D, VandenDriessche T. Biosafety of adenoviral vectors. Curr Gene Ther. (2003) 3:527-543.

17. Chuah MKL, Schiedner G, Thorrez L, Brown B, *et al*. Therapeutic factor VIII levels and negligible toxicity in mouse and dog models of hemophilia A following gene therapy with high-capacity adenoviral vectors. Blood. (2003) 101:1734-1743.

18. Van Damme A, Chuah MKL, Collen D, VandenDriessche T, *et al*. Onco-retroviral and lentiviral vector-based gene therapy for hemophilia: preclinical studies. Semin Thromb Hemost. (2004) 30:185-195.

19. Powell JS, Ragni MV, White GC, Lusher JM, *et al*. Phase 1 trial of FVIII gene transfer for severe hemophilia A using a retroviral construct administered by peripheral intravenous infusion. Blood. (2003) 102:2038-2045.

20. Chuah MK, Brems H, Vanslembrouck V, Collen D, *et al*. Bone marrow stromal cells as targets for gene therapy of hemophilia A. Hum Gene Ther. (1998) 9:353-365.

21. Matsui H. Endothelial progenitor cell-based therapy for hemophilia A. Int J Hematol. (2012) 95:119-124.

22. Ramezani A, Zweier-Renn LA, Hawley RG. Factor VIII delivered by haematopoietic stem cell-derived B cells corrects the phenotype of haemophilia A mice. Thromb Haemost. (2011) 105:676-687.

23. Roth DA, Tawa NE, O'Brien JM, Treco DA, *et al*. Nonviral transfer of the gene encoding coagulation factor VIII in patients with severe hemophilia A. N Engl J Med. (2001) 344:1735-1742.

24. Huang L, Wong YP, Gu H, Cai YJ, *et al*. Stem cell-like properties of human umbilical cord lining epithelial cells and the potential for epidermal reconstitution. Cytotherapy. (2011) 13:145-155.

25. Sivalingam J, Krishnan S, Ng WH, Lee SS, *et al*. Biosafety assessment of site-directed transgene integration in human umbilical cord-lining cells. Mol Ther. (2010) 18:1346-1356.

26. Lynch CM, Israel DI, Kaufman RJ, Miller AD, *et al*. Sequences in the coding region of clotting factor VIII act as dominant inhibitors of RNA accumulation and protein production. Hum Gene Ther. (1993) 4:259-272.

27. Hoeben RC, Fallaux FJ, Cramer SJ, van den Wollenberg DJ, *et al*. Expression of the blood-clotting factor-VIII cDNA is repressed by a transcriptional silencer located in its coding region. Blood. (1995) 85:2447-2454.

28. Tagliavacca L, Wang Q, Kaufman RJ. ATP-dependent dissociation of non-disulfide-linked aggregates of coagulation factor VIII is a rate-limiting step for secretion. Biochemistry. (2000) 39:1973-1981.

29. Lenting PJ, van Mourik JA, Mertens K. The life cycle of coagulation factor VIII in view of its structure and function. Blood. (1998) 92:3983-3996.

30. Plantier JL, Guillet B, Ducasse C, Enjolras N, *et al*. B-domain deleted factor VIII is aggregated and degraded through proteasomal and lysosomal pathways. Thromb Haemost. (2005) 93:824-832.

31. Miao HZ, Sirachainan N, Palmer L, Kucab P, *et al*. Bioengineering of coagulation factor VIII for improved secretion. Blood. (2004) 103:3412-3419.

32. Doering CB, Healey JF, Parker ET, Barrow RT, *et al*. Identification of porcine coagulation factor VIII domains responsible for high level expression via enhanced secretion. J Biol Chem. (2004) 279:6546-6552.

33. Brown HC, Gangadharan B, Doering CB. Enhanced biosynthesis of coagulation factor VIII through diminished engagement of the unfolded protein response. J Biol Chem. (2011) 286:24451-24457.

34. Toschi V. OBI-1, porcine recombinant Factor VIII for the potential treatment of patients with congenital hemophilia A and alloantibodies against human Factor VIII. Curr Opin Mol Ther. (2010) 12:617-625.

35. Healey JF, Parker ET, Barrow RT, Langley TJ, *et al*. The comparative immunogenicity of human and porcine factor VIII in haemophilia A mice. Thromb Haemost. (2009) 102:35-41.

36. Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, *et al*. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Mol Cell Biol. (2001) 21:3926-3934.

37. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, *et al*. Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol. (2006) 357:28-48.

38. Ortiz-Urda S, Thyagarajan B, Keene DR, Lin Q, *et al*. Stable nonviral genetic correction of inherited human skin disease. Nat Med. (2002) 8:1166-1170.

39. Keravala A, Lee S, Thyagarajan B, Olivares EC*, et al*. Mutational derivatives of PhiC31 integrase with increased efficiency and specificity. Mol Ther. (2009) 17:112-120.

40. Sclimenti CR, Thyagarajan B, Calos MP. Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res. (2001) 29:5044-5051.

41. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP*, et al*. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science. (2003) 302:415-419.

42. Recchia A, Bonini C, Magnani Z, Urbinati F*, et al*. Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. Proc Natl Acad Sci U S A. (2006) 103:1457-1462.

43. Wu X, Li Y, Crise B, Burgess SM*, et al*. Transcription start regions in the human genome are favored targets for MLV integration. Science. (2003) 300:1749-1751.

44. Design and performance of the genechip human genome U133 plus 2. 2008.

45. Futreal PA, Coin L, Marshall M, Down T*, et al*. A census of human cancer genes. Nat Rev Cancer. (2004) 4:177-183.

46. Ehrhardt A, Engler JA, Xu H, Cherry AM*, et al*. Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. Hum Gene Ther. (2006) 17:1077-1094.

47. Liu J, Jeppesen I, Nielsen K, Jensen TG*, et al*. Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther. (2006) 13:1188-1190.

48. Liu J, Skjørringe T, Gjetting T, Jensen TG*, et al*. PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. BMC Biotechnol. (2009) 9:31.

49. Giardino D, Corti C, Ballarati L, Colombo D*, et al*. De novo balanced chromosome rearrangements in prenatal diagnosis. Prenat Diagn. (2009) 29:257-265.

50. Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. Am J Hum Genet. (1991) 49:995-1013.

51. Varella-Garcia M, Chen L, Powell RL, Hirsch FR*, et al*. Spectral karyotyping detects chromosome damage in bronchial cells of smokers and patients with cancer. Am J Respir Crit Care Med. (2007) 176:505-512.

52. Fan YS, Siu VM, Jung JH, Xu J*, et al*. Sensitivity of multiple color spectral karyotyping in detecting small interchromosomal rearrangements. Genet Test. (2000) 4:9-14.

53. Hahn WC, Weinberg RA. Rules for making human tumor cells. N Engl J Med. (2002) 347:1593-1603.

54. Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. Nat Genet. (2004) 36:331-334.

55. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. Nature. (2004) 432:307-315.

56. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ*, et al*. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. (2003) 100:3983-3988.

57. Li C, Heidt DG, Dalerba P, Burant CF*, et al*. Identification of pancreatic cancer stem cells. Cancer Res. (2007) 67:1030-1037.

58. Zhou Y, Gan SU, Lin G, Lim YT*, et al*. Characterization of human umbilical cord lining-derived epithelial cells and transplantation potential. Cell Transplant. (2011) 20:1827-1841.

59. Dwarki VJ, Belloni P, Nijjar T, Smith J, *et al*. Gene therapy for hemophilia A: production of therapeutic levels of human factor VIII in vivo in mice. Proc Natl Acad Sci U S A. (1995) 92:1023-1027.

60. Larcher F, Dellambra E, Rico L, Bondanza S, *et al*. Long-term engraftment of single genetically modified human epidermal holoclones enables safety pre-assessment of cutaneous gene therapy. Mol Ther. (2007) 15:1670-1676.

61. Mehrle S, Rohde V, Schlehofer JR. Evidence of chromosomal integration of AAV DNA in human testis tissue. Virus Genes. (2004) 28:61-69.

62. Smith JR, Maguire S, Davis LA, Alexander M, *et al*. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. Stem Cells. (2008) 26:496-504.

63. Hockemeyer D, Soldner F, Beard C, Gao Q, *et al*. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. (2009) 27:851-857.

64. Mulder J, Ariaens A, van den Boomen D, Moolenaar WH, *et al*. p116Rip targets myosin phosphatase to the actin cytoskeleton and is essential for RhoA/ROCK-regulated neuritogenesis. Mol Biol Cell. (2004) 15:5516-5527.

65. Ogata T, Kozuka T, Kanda T. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. J Virol. (2003) 77:9000-9007.

66. Zou J, Sweeney CL, Chou BK, Choi U, *et al*. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. Blood. (2011) 117:5561-5572.

67. Segal DJ, Beerli RR, Blancafort P, Dreier B, *et al*. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry. (2003) 42:2137-2148.

68. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, *et al*. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell. (2008) 31:294-301.

69. Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, *et al*. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods. (2011) 8:67-69.

70. Doyon Y, Vo TD, Mendel MC, Greenberg SG, *et al*. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. (2011) 8:74-79.

71. Guo J, Gaj T, Barbas CF. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol. (2010) 400:96-107.

72. Doyon Y, Choi VM, Xia DF, Vo TD, *et al*. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. Nat Methods. (2010) 7:459-460.

73. Ramakrishna S, Kim YH, Kim H. Stability of zinc finger nuclease protein is enhanced by the proteasome inhibitor MG132. PloS one. (2013) 8:e54282.

74. van Rensburg R, Beyer I, Yao XY, Wang H, *et al*. Chromatin structure of two genomic sites for targeted transgene integration in induced pluripotent stem cells and hematopoietic stem cells. Gene Ther. (2013) 20:201-214.

75. Daboussi F, Zaslavskiy M, Poirot L, Loperfido M, *et al*. Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic Acids Res. (2012) 40:6367-6379.

76. Hartlerode AJ, Scully R. Mechanisms of double-strand break repair in somatic mammalian cells. Biochem J. (2009) 423:157-168.

77. Takashima Y, Sakuraba M, Koizumi T, Sakamoto H, *et al*. Dependence of DNA double strand break repair pathways on cell cycle phase in human lymphoblastoid cells. Environ Mol Mutagen. (2009) 50:815-822.

78. Olsen PA, Gelazauskaite M, Randøl M, Krauss S, *et al*. Analysis of illegitimate genomic integration mediated by zinc-finger nucleases: implications for specificity of targeted gene correction. BMC Mol Biol. (2010) 11:35.

79. Ramirez CL, Certo MT, Mussolino C, Goodwin MJ, *et al*. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucleic Acids Res. (2012) 40:5560-5568.

80. Wang J, Friedman G, Doyon Y, Wang NS, *et al*. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. Genome Res. (2012) 22:1316-1326.

81. Miller JC, Holmes MC, Wang J, Guschin DY, *et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. (2007) 25:778-785.

82. Ramalingam S, Kandavelou K, Rajenderan R, Chandrasegaran S, *et al*. Creating designed zinc-finger nucleases with minimal cytotoxicity. J Mol Biol. (2011) 405:630-641.

83. Qiu P, Shandilya H, D'Alessio JM, O'Connor K, *et al*. Mutation detection using Surveyor nuclease. Biotechniques. (2004) 36:702-707.

84. Preuß E, Treschow A, Newrzela S, Brücher D, *et al*. TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. Hum Gene Ther. (2010) 21:929-941.

85. Tey SK, Dotti G, Rooney CM, Heslop HE, *et al*. Inducible caspase 9 suicide gene to improve the safety of allodepleted T cells after haploidentical stem cell transplantation. Biol Blood Marrow Transplant. (2007) 13:913-924.

86. Gupta A, Meng X, Zhu LJ, Lawson ND, *et al*. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. (1). Nucleic Acids Res. (2011) 39:381-392.

87. Szczepek M, Brondani V, Büchel J, Serrano L, *et al*. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nat Biotechnol. (2007) 25:786-793.

88. Gaj T, Guo J, Kato Y, Sirk SJ, *et al*. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat Methods. (2012) 9:805-807.

89. Cradick TJ, Ambrosini G, Iseli C, Bucher P, *et al*. ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. BMC Bioinformatics. (2011) 12:152.

90. Gabriel R, Lombardo A, Arens A, Miller JC, *et al*. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol. (2011) 29:816-823.

91. DeKelver RC, Choi VM, Moehle EA, Paschon DE, *et al*. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res. (2010) 20:1133-1142.

92. Lombardo A, Cesana D, Genovese P, Di Stefano B, *et al*. Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat Methods. (2011) 8:861-869.

93. Mussolino C, Cathomen T. TALE nucleases: tailored genome engineering made easy. Curr Opin Biotechnol. (2012) 23:644-650.

94. Mali P, Yang L, Esvelt KM, Aach J, *et al*. RNA-guided human genome engineering via Cas9. Science. (2013) 339:823-826.

95. Chang CJ, Bouhassira EE. Zinc-finger nuclease-mediated correction of α-thalassemia in iPS cells. Blood. (2012) 120:3906-3914.

96. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. (2009) 10:57-63.

97. Wu C, Dunbar CE. Stem cell gene therapy: the risks of insertional mutagenesis and approaches to minimize genotoxicity. Frontiers of medicine. (2011) 5:356-371.

Chapter 4

Materials and Methods

4.1. Materials

4.1.1. Chemicals and reagents

All chemicals and reagents were of analytical or ultrapure grade and cell-culture tested where appropriate. The following suppliers (in square parentheses) were sources of: ampicillin, gancyclovir, kanamycin sulfate, thymidine, apidicolin, mimosine, hydroxyurea, nocodozole, etoposide, monoclonal anti-FLAG M2 antibody, monoclonal anti-β-actin antibody, streptavidin-peroxidase polymer (ultrasensitive), DPX mount solution, Mayer's hematoxylin, RNA sample loading buffer, phenol:chloroform:isoamylalcohol 29:28:1, dimethyl sulfoxide, Iscove's modified Eagle's medium, Dulbecco's modified Eagle's medium (DMEM) and CompoZr® targeted integration kit [Sigma-Aldrich, USA]; phospho-histone H2AX (Ser139) rabbit monoclonal antibody Alexa Fluor®647 conjugate [Cell Signalling Technology, USA]; BigDye® Terminator v3.1 cycle sequencing kit [Applied Biosystems, USA]; Surveyor™ mutation detection kit [Transgenomic Inc., USA]; Coatest® SP4 FVIII kit [Chromogenix, Sweden]; VisuLize™ FVIII antigen kit [Affinity Biologicals™ Inc., Canada]; Phusion® Human Specimen direct PCR kit and DyNAzyme EXT DNA polymerase [Finnzymes, USA]; BD Matrigel™ basement membrane matrix and APC BrdU flow kit [BD Biosciences, USA]; anti-vimentin clone V9 [Zymed® Laboratories Inc., USA]; crystal violet [BDH chemicals, UK]; *Pfu* Ultra high-fidelity DNA polymerase, *Pfu* Turbo DNA polymerase, Max efficiency DH5α competent cells and QuikChange™ Lightning Multi Site-Directed Mutagenesis kit [Stratagene, USA]; RNaseOut™ recombinant ribonuclease inhibitor, SuperScript™ II reverse transcriptase, G418 (Geneticin®), hygromycin, Bioprime®DNA labelling kit, and One Shot®Top10 competent cells [Invitrogen, USA]; shrimp alkaline phosphatase, RQ1 RNase-free DNase, Ultrapure™ DNase-RNase free distilled water, Oligo (dT)$_{15}$ primer and CellTiter 96®Aqueous One solution cell proliferation assay kit [Promega, USA]; North2South®Chemiluminescent hybridization and detection kit, BCA protein assay kit and M-PER®Mammalian protein extraction reagent [Pierce, USA]; Hybond N+ nylon membrane, ECL™ Western blotting analysis system and high performance autoradiography film (Amersham hyperfim™ MP) [Amersham Biosciences, GE Healthcare, USA]; nitrocellulose membrane, tetramethylethylenediamine (TEMED), ammonium persulfate, 30% acrylamide/bis solution and iScript™ Advanced cDNA

synthesis kit [BioRad, USA]; T4 DNA ligase, quick blunting kit and all restriction endonucleases [New England Biolabs Inc., UK]; SKY painting probes [Applied Spectral Imaging, Germany]; Tris-acetate-EDTA (TAE), Tris-borate-EDTA (TBE), 10x phosphate-buffered saline (PBS) and agarose [1st Base, Singapore]; fetal calf serum [Hyclone, USA]; TriPure isolation reagent and Expand Long Range dNTP pack [Roche Applied Science, USA]; all DNA ladders (50 bp, 100 bp and 1 kb) and dNTP mix [MBI Fermentas, USA]. HotStar HiFidelity polymerase kit, all plasmid DNA, RNA and genomic DNA isolation kits and PCR purification kits were from Qiagen (Hilden, Germany).

4.1.2 Plasmids

pEGFP-C1, encoding enhanced green fluorescent protein (EGFP) expressed from CMV promoter and pSEAP2-control, encoding secreted alkaline phosphatase expressed from SV40 promoter were from Clontech, USA. pCR®2.1-TOPO® TA cloning kit, pcDNA3.1™3.1(+) and pBudCE4.1 were from Invitrogen Life Technologies, USA. StrataClone™ blunt PCR cloning vector (pSCB) and pKO Scrambler NTKV 1904, encoding *Herpes simplex* virus thymidine kinase (HSV-TK) expressed from MC1 promoter were from Stratagene, USA. pVitro2-mcs was from Invivogen, USA. pSP64-VIII, encoding full-length human factor VIII cDNA, was from the American Type Culture Collection (ATCC), USA. pCMV-Int[1], encoding phiC31 integrase expressed from CMV promoter and pTA-attB, bearing the 300 bp attB fragment, were gifts from Michele P. Calos (Stanford University, USA). MP71-tCD34 –TK.007[2], encoding codon-optimized TK007 was a gift from Boris Fehse (University Medical Centre Hamburg-Eppendorf, Germany). pST1374, a mammalian expression vector for *Fok*I (expressed from CMV promoter), pAAVS1 SA-2A-puro-pA donor [3] (Addgene plasmid 22075), a promoterless puromycin resistance gene expression vector with splice acceptor (SA) site and self-cleaving 2A peptide sequence, pAAVS-CAGGS-EGFP[3] (Addgene plasmid 22212), an expression vector similar to pAAVS1 SA-2A-puro-pA donor with an additional EGFP cDNA expressed from chicken actin promoter, and zinc finger consortium expression vector kit v1.0 were purchased from Addgene (USA). The key plasmids used in this study are depicted in **Appendix 2**.

4.1.3 Primers and oligonucleotides

All primers and oligonucleotides (synthesized at 1st BASE Pte. Ltd., Singapore) used in this study were designed using the following online primer

designing tools: Primer3 (v.0.40) (http://frodo.wi.mit.edu/primer3/) and Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Oligonucleotide sequences are listed in the body of the thesis and **Appendix 1**.

4.1.4 Cell-lines and primary cells

All cell and tissue cultures used in this study were grown in cell culture grade dishes and flasks (Corning®, USA) incubated at 37°C (unless otherwise stated) with 5% CO_2. Unless otherwise stated, cells were cultured in the indicated media supplemented with 10% fetal bovine serum (FBS) (Hyclone Laboratories, USA) in the presence of penicillin 100 (IU/ml) and streptomycin (100 μg/ml) (Gibco®, USA), with media change every three days. Adherent primary human cell types were detached with Accutase™ (eBiosciences, USA) at 37°C for 5 minutes.

K562 (ATCC# CCL-243), a human chronic myelogenous leukemia cell line, Hs746T (ATCC# HTB-135), a gastric carcinoma cell-line and Hs68 (ATCC# CRL-1635, Hs68), normal human foreskin fibroblast cells were purchased from ATCC. K562 was cultured in Iscove's modified Eagle's medium (Sigma-Aldrich) supplemented with 10% FBS while Hs746T and Hs68 were cultured in Dulbecco's modified Eagle's medium (DMEM-25 mM glucose; Sigma-Aldrich) supplemented with 10% FBS. Poietics™ Human mesenchymal stem cells (#18183) purchased from Lonza Biosciences (Singapore) were cultured in mesenchymal stem cell basal medium containing growth supplements (SingleQuots®, Lonza Biosciences).

Primary human dermal fibroblasts (KF1 and NF123), human adipose-derived stromal cells (ADSCs), human bone marrow-derived stromal cells (BMSCs) and CLECs were provided by CellResearch Corporation, Singapore. Dermal fibroblasts, ADSCs and BMSCs were cultured in DMEM-25 mM glucose (Sigma-Aldrich, USA) supplemented with 10% FBS. CLECs were cultured in Medium 171 (Cascade Biologicals, USA) supplemented with 50 ng/ml insulin-like growth factor-1, 50 ng/ml platelet-derived growth factor-BB, 5 ng/ml transforming growth factor-β1, and 5 ng/ml insulin (all from R&D Systems®, USA).

4.1.5 Animals

C57BL/6J mice (Laboratory Animal Centre, National University of Singapore) and exon 16-disrupted FVIII-deficient mice[4] (gift from H. Kazazian, University of Pennsylvania, Philadelphia, PA) were housed in the animal holding units, National Cancer Centre, Singapore (NCCS). Severe combined immunodeficient, non-obese diabetic mice (NOD-SCID; NOD.CB17-Prkdc[scid]) (Animal Resources Centre, Murdoch, Australia) and NOD-SCID-IL2R gamma null

mice (NSG; NOD.Cg-*Prkdc^scid Il2rg^tm1Wjl*/SzJ) (Jackson Laboratory, USA) were housed in specific pathogen-free animal holding units in NCCS. Mice were housed at 20°-24°C in 12-hour light:dark cycles. All animal handling procedures and experimental protocols were approved by the SingHealth Institutional Animal Care and Use Committee. Unless otherwise stated, eight to ten week old mice were used for all experiments.

4.2 Plasmid construction and mutagenesis

4.2.1 Assembly and mutagenesis of B domain-deleted human FVIII constructs

The precursor full length human FVIIII cDNA sequence including the signal peptide is 7055 bp in length and encodes 2351 amino acids (aa). The mature FVIII peptide, 2332 aa long[5], consists of A1 (aa 1-336), A2 (aa 373-710), B (aa741-1648), A3 (aa 1690-2019), C1 (aa 2020-2172) and C2 (aa 2173 -2332) domains.

We assembled a B domain-deleted human FVIII construct from full length FVIII cDNA sequence (NM_000132.3) previously cloned in the plasmid, pSP64-F8. Nucleotides corresponding to B domain aa 1007 to 1648 (NP_000123.1) were deleted in the final version of the FVIII construct, whilst retaining 266 amino acids of the B domain[6] (with 8 glycosylation sites). Briefly, FVIII sequences were PCR amplified (primer sequences given in parentheses below), using high fidelity *PfuTurbo* DNA polymerase, in two separate segments and individually cloned into pCR®2.1-TOPO® TA vectors. The first segment was a 3 kb fragment encoding the A1, A2 domains and part of the B domain (forward primer 5' tgtagcgctagcatgcaaatag 3'; reverse primer 5' gaataaggcgatatctttagtcaa 3') while the second segment was a 2.1 kb fragment encoding part of the B domain, A3, C1 and C2 domains (forward primer 5' gcaaagcccgggaggactgaa 3'; reverse primer 5' cagtggctcgaggtcagtagaggt 3'). The preceding primer sequences incorporated recognition sites for *Nhe*I, *Eco*RV, *Sma*I and *Xho*I used in cloning. DNA sequencing (Big dye terminator v3.1 cycle sequencing kit, Applied Biosystems) was performed with appropriate primers to confirm the correct FVIII sequences of the amplified nucleotide fragments. The two segments were then ligated to generate a single fragment in pCR®2.1-TOPO® TA (blunt ligation of *Eco*RV and *Sma*I/*Xho*I digested fragments) before transferring to pcDNA3.1 (+) (Invitrogen) bearing the CMV promoter (*via Nhe*I and *Xho*I sites). F309S substitution[7] in the A1 domain was performed by site-directed mutagenesis using mutagenic primers (forward 5' agtttctactgtcttgtcatatctct 3' ; reverse 5' agagatatgacaagacagtagaaact 3'), *PfuTurbo* DNA polymerase and *Dpn*I, according to the manufacturer's protocol. The single point mutation was confirmed by sequencing

with the sequencing primer (5' gtcttcatgctgttggtg 3'). This construct was designated pCMV BDD-human FVIII F309S.

4.2.2 Assembly of attB bearing constructs

A 350 bp attB fragment was PCR amplified from pTA-attB to incorporate *Mlu*I ends and cloned into pCR®2.1-TOPO® TA. The attB fragment was transferred to the *Mlu*I site of pEGFP-C1 and designated pEGFP-C1 attB.

The attB fragment was cloned upstream of the CMV promoter in pcDNA3.1 (+) at the *Bgl*II site to derive pattB pCDNA3.1 (+). We assembled pattB CMV BDD-human FVIII F309S by ligating the *Nhe*I/*Xho*I digested 5.1 kb BDD-human FVIII F309S fragment (from pCMV BDD-human FVIII F309S) downstream of the CMV promoter in pattB pCDNA3.1 (+).

A 1.1 kb EF1α promoter fragment was PCR amplified from pBudCE4.1, cloned into pCR®2.1-TOPO® TA and subsequently replaced the CMV promoter of pattB pCDNA3.1 at *Bgl*II and *Nhe*I sites. pattB EF1α BDD–human FVIII F309S was derived by ligating *Nhe*I/*Xho*I digested 5.1 kb BDD-human FVIII F309S fragment downstream of the EF1α promoter.

pattB SV40 BDD-human FVIII F309S was assembled by replacing EF1α promoter in the pattB EF1α BDD–human FVIII F309S construct *via Mun*I/*Nhe*I with a 209 bp SV40 promoter that had been PCR amplified from pSEAP2-control plasmid and cloned into pCR®2.1-TOPO® TA.

pattB Hfer BDD-human FVIII F309S was assembled by replacing EF1α promoter in the pattB EF1α BDD-human FVIII F309S construct *via Mun*I/*Nhe*I with a 1.6 kb human ferritin promoter (modified to consist of CMV enhancer, human ferritin light chain promoter and 5'UTR of chimpanzee elongation factor 1) that had been PCR amplified from pVitro2 plasmid and cloned into pCR®2.1-TOPO® TA.

4.2.3 Assembly of hybrid FVIII cDNA constructs

The precursor full length porcine factor VIII cDNA sequence (NM_214167.1) including the signal peptide is 6401 bp in length and encodes 2133 amino acids (NP_999332.1), whereas the mature FVIII protein is 2113 aa long. The hybrid porcine-human FVIII consisted of porcine A1 and A3 domains and human signal peptide, A2, B (partial), C1 and C2 domains. Porcine FVIII domain cDNAs were derived by RT-PCR of porcine liver RNA. An overlap PCR strategy was used to construct the hybrid FVIII cDNA. The schematic of the cloning strategy is depicted in **Appendix 3.**

Total RNA was isolated from frozen porcine liver tissues using Trizol RNA extraction protocol. First strand cDNA synthesis was performed with 500ng of total RNA, oligo (dT)$_{15}$ primers and SuperScript™ II reverse transcriptase according to standard protocols. Porcine A1 and A3 domain cDNAs were PCR amplified using *PfuTurbo* DNA polymerase and the following primer pairs: porcine A1 domain (A1 forward 5' atgcaaatagagctctccacctgtttctttctgtgtctttgcgattctgctttagtgccatcaggagatactacctgggcgcagt ggaactgt 3' ; A1 reverse 5' aggatgcttcttggcaactgagcggatttggataaagggaga 3'); porcine A3 domain (A3 forward 5' agctttcagaagagaacccgacac 3'; A3 reverse 5' tcccaggggagtctgacacttcttgctgtacaccaggaaagt 3'). Porcine A1 domain forward and reverse primers were designed to contain additional sequences corresponding to human FVIII signal peptide and 5' region of human A2 domain, respectively. Porcine A3 reverse primer was designed to contain additional sequences corresponding to 5' region of human C1 domain. Human A2 and C1 to C2 domains were amplified from pSP64-F8 plasmid with the following primer pairs: A2 to partial B domain (A2 forward 5' tctcccttatccaaatccgctcagttgccaagaagcatcct 3'; B domain reverse 5' gcggggggctctgattttcatcctc 3'); C1 to C2 domains (C1 forward 5' actttcctggtgtacagcaagaagtgtcagactcccctggga 3'; C2 reverse 5' agtgctagctcagtagaggtcctgtgcc 3'). Human A2 and C1 domain forward primers were designed to contain additional sequences corresponding to 3' regions of porcine A1 and A3 domains, respectively. A second round of overlapping PCR amplification (*Nhe*I forward 5' gccgctagcgatgcaaatagagctctcca 3'; B domain reverse 5' gcggggggctctgattttcatcctc 3') was performed using a mixture of porcine A1 and human A2 domain amplicons as the PCR templates. A separate PCR amplification (A3 forward 5' agctttcagaagagaacccgacac 3'; C2 reverse 5' agtgctagctcagtagaggtcctgtgcc 3') was performed using a mixture of porcine A3 and human C1 to C2 domain amplicons as the PCR templates. Both overlapping PCR products (porcine A1-human A2 and porcine A3-human C2) were cloned into StrataClone™ blunt PCR cloning vector (pSCB) and sequenced with appropriate primers to confirm accuracy of amplified sequences. Porcine A3-human C2 fragment excised with *Nhe*I and *Eco*RV and blunt ended using NEB quick blunting kit was ligated with pSCB-porcine A1-human A2 construct at the *Eco*RV site. The full length hybrid FVIII was subsequently digested with *Nhe*I, blunted using NEB quick blunting kit and cloned by blunt ligation to replace BDD-human FVIII sequence (removed with *Nhe*I/*Xba*I digestion) in pattB Hfer BDD-human FVIII F309S construct to derive pattB Hfer hybrid FVIII.

4.2.4 Assembly of *Herpes simplex* virus thymidine kinase-bearing constructs

Codon optimized thymidine kinase (TK007) cDNA was excised from MP71-tCD34 –TK.007 plasmid[2] with *Nco*I and *Hin*dIII digest, blunt-ended using NEB Quick blunting kit and ligated to *Pme*I digested pCDNA3.1 (+) plasmid to derive pCDNA3.1 HSV TK.007.

MC1 promoter was PCR amplified (forward 5' gtcgcgagtcgagcagtgtggtt 3'; reverse 5' ggctagcacgcgcttctacaag 3') from pKO Scrambler NTKV 1904 and cloned into pCR®2.1-TOPO® TA. The cloned MC1 promoter was inserted at *Nru*I and *Nhe* sites into pCDNA3.1 HSV TK.007 to replace the CMV promoter and to derive pMC1 HSV TK.007.

The entire cassette comprising MC1 promoter, HSV TK.007 cDNA and BGH poly A signal sequences was excised with *Nru*I and *Xmn*I and cloned into *Psi*I and *Nae*I digested pZDonor-AAVS1 (CompoZr® targeted integration kit) to derive pZDonor-AAVS1-HSV TK.007.

4.2.5   Assembly and mutagenesis of AAVS1 ZFN constructs

Commercially purchased pST1374 plasmid encodes the catalytic domain of the *Fok*I endonuclease driven from a CMV promoter. Zinc finger peptides cloned into this plasmid are expressed as a fusion protein together with the *Fok*I catalytic domain. *Fok*I is only functional as a dimer and has previously been modified to function as an obligate heterodimer[8], in order to minimize homodimerization which is likely to increase off-target cleavage.

We introduced two point mutations to each *Fok*I monomer as previously described by Miller J.C. *et al.*[8], using PCR mutagenesis. The *Fok*I nuclease to be fused to the right AAVS1-specific ZF was sequentially mutagenised (with mutagenic primers indicated in parenthesis) to incorporate the following changes to the peptide sequence; E490K (forward primer 5' gcaacgatatgtcaaagaaaatcaaacacg 3'; reverse primer 5' cgtgtttgattttctttgacatatcgttgc 3') and I538K (forward primer 5' cacgattaaatcataagactaattgtaatggagc 3'; reverse primer 5' gctccattacaattagtcttatgatttaatcgtg 3') and was designated E490K; I535K *Fok*I. Similarly, the *Fok*I nuclease to be fused to the left AAVS1-specific ZF was sequentially mutagenised (with mutagenic primers indicated in parenthesis) to incorporate the following changes to the peptide sequence; Q468E (forward primer 5' ccaagcagatgaaatggaacgatatgtcgaag 3'; reverse primer 5' ctt cgacatatcgttccatttcatctgcttgg 3') and I499L (forward primer 5' cacgaaacaaacatctcaaccctaatgaatgg 3'; reverse primer 5' ccattcattagggttgagatgtttgtttcgtg 3') and was designated as Q468E; I499L *Fok*I.

DNA encoding a pair of zinc finger peptides specific for the AAVS1 locus[3] were codon-optimized and commercially synthesized (DNA2.0, USA) and cloned into *Xba*I/*Bam*HI digested E490K; I535K *Fok*I and Q468E; I499L *Fok*I plasmids to derive the following constructs, AAVS1 right E490K; I535K *Fok*I and AAVS1 left Q468E; I499L *Fok*I. The entire AAVS1 left Q468E: I499L *Fok*I cassette comprising CMV IE promoter, AAVS1 left zinc finger, Q468E:I499L *Fok*I and BGH poly A signal sequence was PCR amplified using *PfuUltra* high fidelity DNA polymerase to incorporate *Xho*I and *Xba*I ends and cloned into a StrataClone™ blunt PCR cloning vector (pSCB); the resulting construct was designated pSCB AAVS1 left Q468E: I499L*Fok*I. The AAVS1 left ZFN cassette was digested from pSCB AAVS1 left Q468E: I499L *Fok*I with *Xho*I and *Xba*I and cloned into AAVS1 right E490K; I535K *Fok*I plasmid that had been digested with *Sma*I and subsequently dephosphorylated with shrimp alkaline phosphatase (sALP). The resulting dual construct expressing AAVS1 right E490K; I535K *Fok*I and AAVS1 left Q468E; I499L *Fok*I was hereafter known as obligate heterodimer (OH) ZFN.

Guo J *et al.*[9] had previously reported enhanced cleavage activity from a *Fok*I nuclease variant having the following amino acid substitutions, S418P and K441E. We incorporated the S418P (forward primer 5' attgaaattgccagaaatcccactcaggatagaattctt 3'; reverse primer 5' aagaattctatcctgagtgggatttctggcaatttcaat 3') and K441E (forward primer 5' gtttatggatatagaggtgaacatttgggtggatcaagg 3'; reverse primer 5' ccttgatccacccaaatgttcacctctatatccataaac 3') to introduce both substitutions into the AAVS1 right E490K; I535K *Fok*I and pSCB AAVS1 left Q468E: I499L *Fok*I constructs. The mutagenised constructs were designated AAVS1 right Sharkey E490K; I535K *Fok*I and pSCB AAVS1 left Sharkey Q468E; I499L *Fok*I. The AAVS1 left Sharkey ZFN cassette was next digested from pSCB AAVS1 left Sharkey Q468E: I499L *Fok*I with *Xho*I and *Xba*I and cloned into AAVS1 right Sharkey E490K; I535K *Fok*I plasmid that had been *Sma*I digested and dephosphorylated with sALP. The resulting dual expression plasmid was hereafter known as Sharkey ZFN.

A second publication reported further improvements to *Fok*I activity by engineering the catalytic domain. Doyon *et al.*[10] showed that single amino acid changes to each monomeric nuclease domain could restore cleavage activity that is reduced in the obligate heterodimeric forms of *Fok*I. Accordingly, we modified the AAVS1 right Sharkey E490K; I535K *Fok*I construct (forward primer 5' cagcttacacgattaaatcgt aagactaattgtaatgga 3'; reverse primer 5'

tccattacaattagtcttacgatttaatcgtgtaagctg 3') to introduce a H537R amino acid substitution and the pSCB AAVS1 left Sharkey Q468E; I499L *Fok*I construct (forward primer 5' gaagaaaatcaaacacgagacaaacatctcaaccctaat 3'; reverse primer 5' attagggttgagatgtttgtctcgtgtttgattttcttc 3') to create a N496D substitution. These modified constructs were designated AAVS1 right Sharkey E490K; I535K; H537R *Fok*I and pSCB AAVS1 left Sharkey Q468E; I499L;N496D *Fok*I. The AAVS1 left Sharkey ZFN cassette was excised from pSCB AAVS1 left Sharkey Q468E; I499L;N496D *Fok*I with *Xho*I and *Xba*I and ligated into AAVS1 right Sharkey E490K; I535K; H537R *Fok*I plasmid that had been *Sma*I digested and dephosphorylated with sALP. This final ZFN construct was hereafter known as Enhanced Sharkey ZFN.

4.2.6 Assembly of donor constructs for AAVS1 ZFN work

The pZDonor-AAVS1 construct provided by the CompoZr®Targeted Integration Kit – AAVS1 (Sigma-Aldrich) contains 1500 bp homologous to the genomic region spanning the AAVS1 target site, bisected by a 50 bp multiple cloning site into which DNA sequences or genes of interest can be cloned. DNA fragments cloned between the homology arms are integrated into the AAVS1 site during homologous recombination mediated repair following ZFN induced DNA breaks in the genome.

A stuffer fragment encoding recognition sites for *Nde*I, *Mfe*I, *Pac*I and *Bst*Z17I was created by denaturing equimolar mixtures (10 μM each) of 5'phosphorylated oligonucleotides (forward 5' agcttcatatgcaattgttaattaagtatacccaccgagacc 3'; reverse 5' tcgaggtctcggtgggtatacttaattaacaattgcatatga 3') for 10 min at 95°C, followed by annealing at room temperature for 10 min. The stuffer fragment was inserted into the pZDonor construct digested with *Hin*dIII and *Xho*I to derive pZDonor-AAVS1 stuffer.

An entire 3.75 kb DNA fragment (consisting of CMV promoter, EGFP, SV40 poly A, SV40 promoter, kanamycin/neomycin resistance gene and HSV TK poly A) was excised from pEGFP-C1 with *Ase*I and *Bsa*I, and cloned into *Nde*I/*Bsa*I digested pZDonor-AAVS1 stuffer construct to derive pZDonor EGFP.

Similarly, pZDonor hybrid FVIII construct was derived by cloning a 9.123 kb *Mfe*I/*Bst*Z17I digested fragment (consisting of CMV enhancer, human ferritin light chain promoter, hybrid FVIII cDNA, BGH poly A, neomycin resistance gene, SV40 poly A) from pattB Hfer hybrid FVIII into *Mfe*I/*Bst*Z17I digested pZDonor-AAVS1 stuffer.

The 9.123 kb *Mfe*I/*Bst*Z17*I* digested hybrid FVIII mentioned in the preceding paragraph was also cloned into *Mfe*I/*Bst*Z17I pZDonor-AAVS1-HSV TK.007 plasmid to derive the construct pZDonor hybrid FVIII TK.007.

Lastly, pSA-2A-Puro-Hybrid FVIII construct was cloned by ligating a blunt-ended, 7.877 kb *Mfe*I/*Avr*II digested fragment (consisting of CMV enhancer, human ferritin light chain promoter, hybrid FVIII cDNA, BGH poly A) from pattB Hfer hybrid FVIII into *Sal*I digested, blunt-ended and dephosphorylated pSA-2A-Puro-pA donor plasmid.

4.3. PhiC31 integrase modification of CLECs

4.3.1. Isolation, culture and characterization of CLECs

The isolation and culture of CLECs has been described previously [11-14] (International publication number W0 2006/019357A1 and UK patent GB2432166). Briefly, fresh umbilical cords from uncomplicated pregnancies were transported in L-15 medium supplemented with 50 IU/ml penicillin, 50 μg/ml streptomycin, 250 μg/ml amphotericin B (Fungizone) and 50 μg/ml gentamicin, and processed in sterile conditions. Blood was removed by flushing the cannulated cord with phosphate-buffered saline (PBS) supplemented with 5 IU/ml heparin (Sigma-Aldrich). The cord was next cut into 2 cm segments, washed with PBS, disinfected with 70% ethanol and washed again with antibiotic-containing PBS. The amniotic membrane was dissected free from other cord contents, cut into 0.5 cm^2 squares and placed in a cell culture dish filled with 5 ml of Medium 171 (Cascade Biologics). Explants were cultured at 37 °C/5% CO2, with media change every 3 days. Outgrowing cells were trypsinized (0.0125% trypsin/0.05% EDTA) and seeded at a density of 1×10^6 cells/dish in complete medium as detailed in section 4.1.4. Cells were subcultured at 70% confluency and expanded or cryopreserved.

CLECs were analyzed by reverse transcription (RT)-PCR and protein immunoblotting for expression of pluripotency markers. A human embryonic stem cell line (HUES) and human primary dermal fibroblast cells served as positive and negative controls, respectively, for these characterization experiments. RNA extracted from cells (RNeasy mini kit; Qiagen) was treated with DNAse I (MBI Fermentas), reverse transcribed (SuperScript™ II reverse transcriptase) and amplified using GoTaq qPCR master mix (Promega) and the following PCR primers: *Nanog* (forward primer 5′ ttccttcctccatggatctg 3′; reverse primer 5′ tctgctggaggctgaggtat 3′), *Oct-4* (forward primer 5′ ggttctatttgggaaggtattcag 3′; reverse primer 5′ ggtttctgctttgcatatctc 3′) and γ-actin (forward primer 5′ accactggcattgtcatggactct 3′; reverse primer 5′

305

atcttgatcttcatggtgctgggc 3′). Amplified products were electrophoresed on 2% agarose gels, imaged using GelDoc 2000 transilluminator (Bio-Rad Laboratories, Hercules, CA) and quantified by densitometry using QuantityOne software.

For protein immunoblot analysis, cells were lysed with M-PER mammalian protein extraction reagent (Pierce, Waltham, MA); 20–50 µg protein from each cell lysate was separated by 14% SDS-PAGE under reducing conditions, electroblotted onto nitrocellulose membrane (Bio-Rad Laboratories) and probed with specific antibodies against human *Oct-4* and *Nanog* (sc-5279; Santa Cruz Biotechnology, Santa Cruz, CA and ab21624; Abcam, Cambridge, UK, respectively). Antibody binding was visualized by horseradish peroxidase-conjugated goat anti-mouse or goat anti-rabbit secondary antibodies (Promega and Santa Cruz Biotechnology, respectively) and a chemiluminescence-based photoblot system (Amersham Biosciences, Piscataway, NJ).

4.3.2. Gene transfer

Gene transfer studies for the phiC31 integrase work were performed mainly using the BTX electroporation system. For optimization studies, 2 million CLECs in 400 µl of solution NC [15] were electroporated with 5 µg of pEGFP-C1 plasmid DNA in a 0.4 cm cuvette. Electrotransfer was performed with a single pulse delivered by BTX ECM 830 electroporator (Genetronics, San Diego, CA). For voltage optimization, pulse duration was fixed at 25 ms at different voltages from 160 V to 280 V. For pulse duration optimization, the voltage was maintained at 240 V while the pulse duration varied from 10 to 40 ms. For evaluating different electroporation buffers (RPMI 1640/10% FCS, Mirus solution, solution NC and OptiMEM), the voltage and pulse duration were fixed at 240 V and 30 ms, respectively.

For phiC31 integrase work, CLECs were co-transfected by electroporating 2 × 10⁶ cells with 10 µg pattB Hfer BDD-human FVIII F309S or pEGFP-C1 attB and 1.5 µg of pCMV-Int in 400 µl of RPMI 1640/10% FCS medium with a single pulse delivered at 165V and 55 ms. Mixed populations of stably integrated CLECs were selected by culture in medium containing 0.6 mg/ml G418 for 7 days.

4.3.2.1. Transfection efficiency

Observation and imaging of GFP-positive cells on culture flasks was done using a Nikon TE-300 inverted fluorescence microscope equipped with a filter for FITC (excitation= 425nm; emission= 500nm) and Nikon ACT-1 software.

Electroporated cells were detached from culture flasks by incubating with Accutase™ solution at 37°C for 5 min and spun down at 1500 rpm for 5 min

(Eppendorf 5810R centrifuge) to obtain a cell pellet that was resuspended in 500 μl of PBS containing 1μg/ml propidium iodide and transferred to 5 ml round bottom BD Falcon™ polystyrene tubes (FACS tubes) (BD Biosciences, USA). The percentages of GFP-positive cells and propidium iodide-positive cells were determined by flow cytometry (BD FACSCalibur; BD Biosciences, USA) using a 488 nm argon ion laser for excitation and 530/30 nm bandpass filter in the FL1 channel for GFP detection and 585/42 nm bandpass filter in the FL2 channel for propidium iodide detection. Results were analyzed using FlowJo v7.22 software to determine percentage of cells positive for GFP (transfection efficiency) and propidium iodide (percent mortality).

4.3.2.2. Integration frequency

The percentage of cells integrated pattBGFP following treatment with phiC31 integrase was estimated from the number of stable integrants from seeding FACS-sorted EGFP-expressing CLECs (2,000 and 5,000 cells) into 10 cm petri dishes (done in triplicate) followed by G418 selection. The number of GFP+ cells/clones remaining after 7 days of selection was scored manually by visualizing under a fluorescence microscope to obtain the average integration frequency.

4.3.3. Factor VIII measurements

FVIII levels in conditioned media were determined using a chromogenic assay which measures FVIII activity, while an ELISA-based method was used to quantify human-specific FVIII antigen levels in murine plasma.

4.3.3.1. Chromogenic FVIII assay

The principle of the chromogenic FVIII activity assay is as follows. The rate of conversion of factor X to factor Xa, in the presence of excess factor IXa, calcium and phospholipid, is directly proportional to the activity of the cofactor, factor VIIIa. Factor Xa hydrolyses a chromogenic substrate, S-2765, to liberate a chromophoric group, pNA, which is quantified spectrophotometrically at 405 nm. The intensity of the color of released pNA from the hydrolyzed substrate generated by factor Xa is proportional to factor VIII activity in the sample.

Factor VIII levels were measured using a Chromogenix Coamatic®Factor VIII kit (Chromogenix, Sweden) according to manufacturer's protocol. Conditioned media were either used directly or frozen at -80°C until assayed. Generally, a standard curve was obtained by diluting factor VIII standards (Dade Behering, Germany) to concentrations ranging from 1 IU/ml to 0.01 IU/ml. The various standards and samples were diluted 1:80 (or at a higher dilution when necessary) and

assayed in a Costar® 96-well flat-bottom microplate (Corning®, USA). End point absorbance readings at 405 nm were measured using a Dynex MRX II plate reader (Dynex Technologies Inc., USA). FVIII levels of test samples **were determined from the** standard curve in which levels of FVIII standards (IU/ml) were plotted against absorbance readings at 405 nm.

4.3.3.2. ELISA FVIII assay

Visulize™ FVIII antigen ELISA kit (Affinity Biologicals, Canada) consists of a 96-well microplate coated with sheep polyclonal antibody against human FVIII. Antibody-bound human FVIII antigen in test samples is detected by peroxidase-labeled anti-FVIII antibodies that convert the peroxidase substrate tetramethylbenzidine (TMB) to a blue colored product that is spectrophotometrically measured at 450 nm. Standards (provided in the kit) and citrated test plasma samples were diluted 1:4 prior to assay according to manufacturer's protocol and absorbance readings at 450 nm were measured using a Dynex MRX® II plate. A standard curve derived by plotting levels of FVIII standards to absorbance readings at 450 nm was used to determine FVIII antigen levels in test samples.

4.3.4. Documenting integration sites

Integration profile of CLECs treated with phiC31 integrase was determined by sequencing donor DNA-genomic DNA fragments retrieved by a plasmid rescue method. Genomic cross-over junctions identified by the presence of attB sequences flanking donor vector sequences were assigned to specific genomic locations by bioinformatic analyses and evaluated for chromosomal abnormalities, such as insertions or deletions, at the integration junctions. We sought to characterize the genomic profile of recovered integration events as described in the following sections.

4.3.4.1. Plasmid rescue

The plasmid rescue method uses a combination of restriction enzymes to fragment genomic DNA without digesting within the donor vector sequences such that these digested fragments can be circularized by ligation, transformed into bacterial hosts and re-isolated as circular plasmids for sequencing and further characterizations to yield a genomic snapshot of integration events (**Appendix 4A**).

Genomic DNA from phiC31 integrase modified CLECs (either from clonal or pooled populations of cells) was isolated using Qiagen Blood and Cell Culture Miniprep Kit and digested overnight with a combination of either *Spe*I, *Xba*I and

*Nhe*I or *Bam*HI and *Bgl*II (10 U of each enzyme per µg of genomic DNA). Digested DNA was extracted with 25:24:1 phenol-chloroform-isoamylalcohol (Sigma-Aldrich), precipitated with one-tenth the volume of 3 M sodium acetate at pH 5 and two volumes of 100% ethanol. Precipitated DNA was washed with 70% ethanol before redissolving in nuclease-free water. Digested DNA was circularized by ligating under dilute conditions (1-10 ng/µl) with 1U of T4 DNA ligase (NEB) at 16°C overnight. Ligated products were ethanol precipitated, cleaned up with phenol-chloroform clean-up, washed with 70% ethanol and redissolved in 10 µl of nuclease-free water before being electroporated (1-5 µl of ligated products) into 50 µl of electrocompetent DH10B *E.coli* (Invitrogen) at 1.85kV, 25µF and 200Ω using a 1 mm cuvette in a Gene Pulser (Bio-Rad Laboratories). Electroporated cells were recovered into 700 µl of SOC medium, incubated at 37°C for an hour with shaking at 900 rpm in a thermal mixer (Eppendorf), pelleted by centrifugation at 7000 rpm (Eppendorf) for 3 min before resuspension in 50 µl of SOC media, plating on Luria-Bertani (LB)-agar plates containing 50 µg/ml kanamycin and cultured overnight at 37°C. Antibiotic-resistant bacterial colonies were picked into 5ml of LB-media containing 50 µg/ml of kanamycin and cultured overnight at 37°C. Plasmids were isolated from bacterial cultures using Qiagen Plasmid Midi kits and quantified using a Nanodrop®ND-1000 Spectrophotometer (Thermo Scientific, USA).

4.3.4.2. Characterization of retrieved integration events

4.3.4.2.1.   Sequencing of rescued plasmids

Rescued plasmids (500 ng) were sequenced with BigDye®Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and a primer (CHOSeq R: 5'tcccgtgctcaccgtgaccac3') which was specific to the attB sequence and used to read into the genomic region adjacent to a cross-over (integration) site[16].

4.3.4.2.2.   Characterizing integration sites

DNA sequences were mapped to the reference human genome (GRCh37/hg19) using the BLAT program (http://genome.ucsc.edu) and cross-over junctions (genomic integration sites) were identified by aligning retrieved sequences to the attB sequence (**Appendix 4B**). The identified integration sites were mapped to chromosome cytobands and characterized by their position within exons, introns, repeat elements or intergenic regions, and distance to the nearest transcription start site. Identified integration sites were also noted for their proximity to potential oncogenes and tumor suppressor genes from a compilation of 1650 possible oncogenes and tumor suppressor genes (http://www.bushmanlab.org/links/genelists).

Further analyses determined if there were insertions or deletions of either genomic DNA or donor vector at the cross-over junctions.

4.3.4.2.3.   Motif search at recovered integration sites

To determine whether the recovered integration sites shared a common motif at the cross-over junction, 100 bp of genomic DNA sequences flanking the cross-over point of integration events were retrieved from the reference human genome sequence (version hg 19) and analyzed using the motif search program Multiple Em for Motif Elicitation (MEME; http://meme.nbcr.net), which detects sequence similarities and generates a common motif (with probability e-values), if present.

4.3.4.2.4.   Screening of CLECs for integration of pattB Hfer hybrid FVIII at 8p22 locus

CLECs electroporated with 12 μg of pattB Hfer hybrid FVIII only or together with 2.5 μg pCMV-Int plasmid using nucleofector primary cell solution P1 (setting CM113) and a nucleofector 4D device (Lonza) were either cultured unselected or selected with 1 mg/ml of G418 for 5 days starting from day 6 post-electroporation. FVIII activity assay (Coamatic®Factor VIII kit, Chromogenix) was performed on overnight conditioned media of genome-modified CLECs seeded at 100 000 cells per well (12-well plate, Nunc) in 500 μl of culture media on days 6 and 25 post-electroporation.

Genomic DNA was extracted from bulk population of genome-modified CLECs using QIAamp DNA mini kit. Junctional PCR was performed on 200 ng of genomic DNA using DyNAzyme EXT DNA polymerase. Left integration junction PCR (454 bp) was performed using primer pairs specific to a locus on chromosome 8p22 (forward: 5' gggctctggagtaaaggtgaaa 3') and donor vector (reverse: 5' gttcgccgggatcaactacc 3'). Right integration junction PCR (333 bp) was performed using primer pairs specific to the vector sequence (forward: 5' tcgacgatgtaggtcacgg 3') and chromosome 8p22 genomic DNA (reverse: 5' gcatggcctcatttccgtct 3'). Control genomic PCR (900 bp) was performed with AAVS1 genomic primers (forward: 5' aagaagcgcaccacctccaggttct 3'; reverse: 5' atgacctcatgctcttggccctcgta 3'). All PCR was set up in a 20 μl reaction volume and amplified for 30 cycles using a PTC-200 Peltier gradient thermal cycler (MJ Research Inc., USA) at an annealing temperature of 56°C and extension time of 1 min per cycle. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and quantified using QuantityOne software.

To screen clonal CLECs for transgene integration at chromosome 8p22 locus[17], G418-resistant stable population of CLECs (pattB Hfer hybrid FVIII + pCMV Int, G418 selected) were flow sorted (4 cells per well) into individual wells of 96-well plates (Nunc) and allowed to expand in culture. A replica plate for further analysis of clones was established by detaching and re-seeding cells into a second 96-well plate. Cells for 8p22 integration screening (in 96-well plates) were lysed *in situ* with 60 µl of lysis buffer and screened by Direct PCR (section 4.4.1.6) for the presence of integration junctions using Phusion Human Specimen Direct PCR kit and the same sets of primers mentioned earlier in this section. Clones identified to be positive for transgene integration at 8p22 locus were retrieved from the replica plate and analyzed for FVIII activity and for transgene copy number by fluorescence *in situ* hybridization (FISH) (section 4.3.7). Further studies were performed on selected clones that had high levels of FVIII expression and single copy transgene integration.

4.3.4.2.5. Reverse transcription and quantitative PCR to determine changes in *DLC1* transcript levels

Reverse transcription (RT)-quantitative PCR was employed to determine changes in *DLC1* transcript levels in clonal CLECs with transgene integration in 8p22. CLECs electroporated without plasmid DNA and of the same number of population doublings were controls.

Total RNA was extracted from approximately 1×10^6 cells using TriPure isolation reagent as detailed in section 4.5.2, treated with DNase I to remove contaminating genomic DNA and purified using RNeasy® Mini kit. Reverse transcription was performed on approximately 1µg of DNAse-treated RNA using iScript™ Advanced cDNA synthesis kit (Bio-Rad Laboratories), according to the manufacturer's instruction. Quantitative-PCR (Q-PCR) was performed using 2 µl of 1st strand cDNA in a 20 µl reaction volume using GoTaq® qPCR Master Mix and 45 cycles at an annealing and extension temperature of 62°C using CFX96™ Real-Time PCR detection system (Bio-Rad Laboratories). Intron-spanning exonic primers were used to amplify *DLC1* exons 1-2 (F: 5'-tcctgccccaatggaatgtc-3'; R: 5'-gttggtgtgcctgatggaga-3'), exons 8-9 (F: 5'-gaaggggatgcagcggatag-3'; R: 5'-agcagggccgttagctttag-3') and a housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) exon 6-7 (F: 5'-gcctcctgcaccaccaact-3'; R: 5'-cgcctgcttcaccaccttc-3'). *DLC1* transcript levels were normalized to *GAPDH* expression levels and the fold-change in *DLC1* transcript levels in clonal CLECs with 8p22 integration was reported relative to *DLC1* transcript levels in control CLECs, using the 'delta-delta C(T) method'[18].

4.3.5. Transcriptional profiling

Total RNA isolated from naive and genome-modified CLECs served as starting material for transcript profiling on Human Genome U133 Plus 2.0 Arrays (Affymetrix, Santa Carla, CA) while those from genome-modified oligoclonal CLECs with integrations at 8p22 were evaluated using GeneChip® PrimeView™ Human Gene Expression array (Affymetrix), following the recommended protocols. Transcription expression data were analyzed using GeneChip Operating Software (Affymetrix). Transcripts whose expression levels differed significantly (determined by Wilcoxon signed-rank test) by more than two-fold in genome-modified CLECs compared to naïve CLECs were considered significantly altered and were further analyzed. DAVID (Database for Annotation, Visualization and Integrated Discovery) 2.1 Functional Annotation Tool (http://david.abcc.ncifcrf.gov) was used to ascribe functions and other annotations for significantly altered transcripts and for pathway mapping.

Genes in which integration sites occurred and genes within a 1MB window centered on integration sites were analyzed to determine for significant alterations in their transcripts levels. Altered transcripts were also referenced to a compilation of known proto-oncogenes and tumor suppressor genes (http://www.bushmanlab.org/links/genelists) in order to identify if they belonged to either category.

4.3.6. Genome copy number change analyses

High-resolution copy number profiling was performed on genomic DNA of naive and genome-modified CLECs using the Human Mapping 500K Array Set (Affymetrix) [data analyzed using GeneChip Chromosome Copy Number Analysis Tool] while oligoclonal genome-modified CLECs with 8p22integrations were evaluated using a higher resolution and newer array, Cytoscan® HD array (Affymetrix) [data analyzed using Chromosome Analysis Suite]. Regions of copy number gain or loss were defined as having ≥ 3 consecutive SNPs or ≥ 50 consecutive probes concordant for significant copy number abnormalities when analyzed using the Human Mapping 500K Array and Cytoscan HD array, respectively. Log_2 signal intensity ratios >0.3 and ≤ -0.3 were criteria for significant copy number gain and loss, respectively.

Integration sites and non-integration sites associated with significantly altered transcripts were compared to copy number profiling data to determine if there were any significant copy number gains or losses at these loci.

4.3.7. Fluorescence *in situ* hybridization

Fluorescence *in situ* hybridization (FISH) to detect integration of pAttB GFP and pAttB Hybrid FVIII vectors were performed on interphase nuclei of genome-modified CLECs with fluorescein-12-dUTP (PerkinElmer, Waltham, MA) labeled probes generated by PCR amplification of EGFP cDNA from pattB EGFP-C1 (740 bp PCR product: forward primer 5' ccggtcgccaccatggtgag ; reverse primer 5' ctgagtccggacttgtacag 3') and PCR amplification of neomycin cDNA from pAttB Hybrid FVIII (817 bp PCR product: forward primer 5' ttgcacgcaggttctccggc 3'; reverse primer 5' ggcgtcgcttggtcggtcat 3'), respectively. For investigation of transgene integration into 8p22, Texas-Red-5-dUTP labeled human chromosome 8 centromeric probes (Children's Hospital Oakland Research Institute) were used. All probes were labeled by random prime labeling (BioPrime DNA Labeling System, Invitrogen), cleaned with ChargeSwitch PCR Clean-up kit (Invitrogen) to remove excess fluorochrome, ethanol precipitated and resuspended in 15 μl of hybridization solution (50% formamide and 10% dextran sulphate in 2 x SSC). Suspensions of CLECs in fixative solution consisting of 3 parts methanol and 1 part acetic acid (v/v) at -20°C, were spread onto clean polylysine-coated glass slides (Thermo Scientific) by the dropping method[19] and the slides were aged overnight at 56°C. Slides were treated with pepsin (12 μl of 100mg/ml pepsin in 0.1 M HCl) at 37°C for 10 min, fixed in 1% formaldehyde for 10 min at room temperature, dehydrated sequentially in 70%, 80% and 100% ethanol (2 min each) and air dried. Slides were incubated at 70°C for 2 min to denature nuclear DNA and subjected to further sequential dehydration in 70%, 80% and 100% ethanol (2 min each) before air drying. Labeled probes were denatured at 72°C for 6 min before adding to slides which were then cover-slipped and sealed with rubber glue prior to hybridization at 37°C in a humidified chamber for 48-72 hours. Hybridized slides were washed sequentially in solutions containing 50% formamide/2xSSC, 2xSSC and 0.1xSSC, air-dried, counterstained with 4,6-diamino-2-phenylindole (DAPI) and enumerated for probe signals by visualizing with an Olympus BX61 epifluorescence microscope (Olympus, Tokyo, Japan).

4.3.8. Karyotype and spectral karyotype

Conventional G-banding karyotype and spectral karyotype (SKY) studies were performed on cells arrested at metaphase by 0.1 μg/ml colcemid (Invitrogen), a microtubule depolymerising mitotic inhibitor drug, for 3 hrs at 37°C. Mitotic cells detach from adherent cell culture and could be harvested, washed with PBS and

incubated in hypotonic 0.06 M KCl at 37°C for 10 min. Cells were pelleted by centrifugation at 900 rpm for 12 min and washed thrice with fixative solution (3 parts methanol and 1 part acetic acid, v/v) before final resuspension and storage in 0.5 ml of fixative at -20°C. Metaphase preparations were spread onto polylysine-coated glass slides by the dropping method[19] and the slides were aged for 3 to 21 days (in a closed container) at room temperature before use.

G-banding was performed on slides containing metaphase preparations aged for at least 1 week. Slides were incubated in 2 x SSC at 65°C for 1-2 hrs, transferred to 0.85% (w/v) NaCl solution at room temperature for 5 min, drained and incubated horizontally in 0.85% NaCl containing 0.025% trypsin for 15 -20 s followed by a rinse in 0.85% NaCl before staining with Giemsa stain (1 part Giemsa stain/3 parts Gurr's phosphate buffer, v/v, pH 6.8) for 2 min. Slides were rinsed with Gurr's buffer, air-dried and mounted with DPX mount solution, and cover-slipped before imaging with an Olympus BX61 epifluorescence microscope and BandView® software (Applied Spectral Imaging, Germany).

Spectral Karyotyping was also performed on slides of metaphase preparations aged for at least 1 week. Slides were pre-treated with 2 x SSC for 2-3 hrs at 37°C, incubated in 0.01 M HCl with 15 µl of pepsin (100 mg/ml) at 37°C for 7 min, washed sequentially twice with PBS and once with PBS/0.05 M MgCl₂ before incubation in 1% formaldehyde at 25ºC for 10 min. Slides were then washed once with PBS, dehydrated sequentially in 70%, 80% and 100% ethanol and air-dried. Chromosomal DNA was denatured by incubating slides at 70°C for 2 min in 70% formamide in 2 x SSC, pH 7.0. Slides were immediately subjected to sequential dehydration in cold 70%, 80% and 100% ethanol (2 min each) before drying in air. SKY paint reagent (vial#1) was denatured at 80°C for 6 min, allowed to re-anneal at 37°C for 1 hr before being adding to slides, which were then cover-slipped and sealed with rubber glue prior to hybridization at 37°C in a humidified chamber for 48 hours. Hybridized slides were washed sequentially in solutions containing 50% formamide/2xSSC, 1xSSC and 4xSSC/0.1% Tween 20 and air-dried. Slides were next blocked with 60 µl of blocking reagent (SKY reagent vial#2) at 37°C for 30 min, rinsed in wash solution III and drained dry. About 60-80 µl of SKY paint reagent [CAD kit vial#3 (Applied Spectral Imaging) diluted as follows; 500 µl of 4xSSC + 2.5 µl anti-Digoxin (Sigma, #D8156) + 2.5 µl Cy5 Strep Avidin(CAD kit vial#3)] was added to slides which were incubated (with cover-slipping) at 37°C for 1hr. After a triple wash in solution III (4xSSC/0.1% Tween20) at 45°C, 60 -80 µl of SKY paint reagent (CAD kit vial#4 (Applied Spectral Imaging) diluted as follows; 500 µl of 4xSSC + 2.5 µl Cy5.5 anti mouse (CAD kit vial#4)]) was added to slides which were incubated at 37°C for 1

hr. Slides were then washed thrice in wash solution III at 45°C, rinsed in distilled water, allowed to air-dry before counterstaining with 4,6-diamino-2-phenylindole (DAPI), cover-slipped and sealed with nail polish. SKY reagent labeled metaphase chromosomes were imaged with an Olympus BX61 epifluorescence microscope and analyzed using BandView® software. A minimum of 40 metaphases were examined for each sample.

4.3.9. Tumorigenicity assessment

Long term *ex vivo* culture or genetic modification of cells may induce cellular transformation characterized by unlimited proliferation *in vitro*, loss of contact inhibition, adherence independence and tumor formation in immunocompromised mice models. Tumorigenicity assessment studies are pertinent to evaluate gene modified cells for acquired oncogenicity.

4.3.9.1. *In vitro* colony formation assay

In vitro colony formation assay is a crude quantitative assessment of the proliferative characteristics of clonal cell populations based on the number and size of colonies formed.

Typically, untreated or stably modified cells were plated at a low seeding density (2000 or 5000 cells) in a 10 cm dish and allowed to grow for 14 days, with medium change every 3 days. Cultures were then washed with PBS, fixed in 10% formalin for 15 min at room temperature, rinsed with PBS and stained with 1% (w/v) crystal violet (BDH Chemicals, UK) for 30 min at room temperature. Following three rinses with PBS and air drying, dishes with stained colonies were imaged with an inverted microscope (Axiovert 25CFC, Carl Zeiss). The number and sizes of colonies were manually scored using KS400 software (Carl Zeiss).

4.3.9.2. *In vivo* tumorigenicity assay

Tumorigenicity potential of cells may be evaluated by implantation into immunocompromised mice models (SCID, Nude, NOD-SCID) and observing for tumor formation.

The tumorigenic potential of genome-modified CLECs was assessed in NOD-SCID mice by subcutaneous nuchal and renal subcapsular implantation of 5×10^6 cells suspended in 50 µl PBS. Mice were visually inspected weekly for the appearance of subcutaneous tumors for 4 months, and killed after 1 and 3 months to check for renal subcapsular tumors. Survival and engraftment of cells at implantation sites were shown by immunohistochemical staining of tissue sections taken from

implantation sites using an antibody directed against a suitable marker for human cells.

4.3.9.3. Immunohistochemistry

Basically, tissues were excised from implantation sites, fixed overnight in 10% formalin, processed through an ethanol series (70%, 80% and 100% ethanol) in a tissue processor (Leica Microsystems, Germany) and embedded in paraffin wax. Five micron sections of paraffin-embedded tissues were prepared using a microtome (Leica Microsystems), placed on microscope slides and baked at 56°C overnight for immunohistochemical staining.

Typically, slides were deparaffinized by two incubations in xylene for 5 min, hydrated by sequential incubation in 100%, 95% and 70% ethanol, rinsed in tap water and blocked with 10% serum (from the host species of secondary antibodies) for 30 min. Tissue sections were rinsed thrice with PBS and after each of the incubation steps mentioned below. Tissues were incubated with 1:3 diluted (DAKO REAL antibody diluent) mouse anti-human vimentin antibody (Clone V9, Zymed) for 1 hr at 37°C followed by 30 min incubation with 1:2 diluted (same diluent as above) DAKO REAL EnVision rabbit/mouse HRP-conjugated secondary antibody. Following thorough rinsing with PBS, DAB chromogen (1 part DAB chromogen + 3 parts DAKO REAL substrate buffer) was added to slides and incubated until a brown color was visible in the tissue sections. Excess chromogen was rinsed off with PBS, slides counterstained with Mayer's hematoxylin for 3min, rinsed under running tap water, dehydrated by sequential incubation in 70%, 95% and 100% ethanol, incubated with xylene before coverslipping with DPX mount solution. Tissue sections were visualized and imaged with an inverted microscope (Axiovert 25CFC, Carl Zeiss, Germany) under 20x magnification, using the KS400 software.

4.3.10. Factor VIII study

CLECs stably integrated with a donor construct expressing BDD-human FVIII F309S cDNA from a human ferritin promoter were selected on the basis of G418 resistance. Culture supernatants were assayed for FVIII activity using the Coatest® SP4 FVIII kit.

4.3.10.1.   Implantation of FVIII-secreting cells

Approximately 8×10^6 CLECs stably secreting FVIII were implanted either unencapsulated or encapsulated with Matrigel™ basement membrane matrix into the subcutaneous nuchal region of anesthetized hemophilic male mice (n=5 for each

group). Control mice were implanted with a similar number of naive, Matrigel™-encapsulated CLECs.

4.3.10.2.   Phenotypic correction/blood loss assay

Blood samples for FVIII assays were collected before, 1, 3, 6 and 15 days after cellular implantation. Blood was obtained by puncturing the retro-orbital venous plexus with heparinised capillary tubes and collected in 0.1 volume of 0.1 mol/L sodium citrate. Plasma was obtained by centrifugation at 20,000xg at 4°C for 10 min and stored at -80°C until the time of assay.

Human FVIII antigen levels in mouse plasma samples were determined using a human FVIII-specific ELISA kit as described in section 4.3.3.2. Assessment of phenotypic correction was performed by determining the volume of blood loss in hemophilic mice following a hemostatic challenge as described previously[20]. Typically, the tail was warmed to 37 °C for 2 min, then severed 2 cm from the tip with a sharp blade and immediately placed in a microfuge tube containing 0.5 ml PBS at 37 °C for 15 min, after which bleeding was arrested by cauterization using a Gemini cautery system kit (Harvard Apparatus, USA). The difference in the weight of the tube before and after blood collection quantified blood loss (expressed as mg of blood loss).

4.4.   AAVS1 ZFN modification of cells

We attempted to evaluate site-specific genomic integration of donor inserts of varying sizes (50 bp, 1 kb, 4 kb and 9 kb) into intron 1 of the *PPP1R12C* gene (AAVS1 site) using AAVS1 ZFNs that were either commercially purchased (Sigma-Aldrich) or with ZFNs that we assembled as detailed in section 4.2.5. Three donor plasmids with neomycin resistance antibiotic selection marker that were used in the initial study to integrate DNA fragments of increasing size (sizes indicated in parenthesis) were:

pZDonor-AAVS1 (50 bp multiple cloning site)

pZDonor GFP (3.75 kb fragment consisting of CMV promoter, EGFP, SV40 poly A, SV40 promoter, kanamycin/neomycin resistance gene and HSV TK poly A)

pZDonor hybrid FVIII (9.123 kb fragment consisting of CMV enhancer, human ferritin light chain promoter, hybrid FVIII, BGH poly A, neomycin resistance gene, SV40 poly A).

Initial optimization studies were performed on a leukemia cell line, K562, while subsequent studies focused on different primary human cell types.

Subsequent studies were performed with donor vectors with a promoterless puromycin selection cassette, splice-acceptor sequence and self-cleaving 2A-peptide sequence[3], which resulted in more efficient selection of cells with targeted transgene integration. Commercially purchased pAAVS1 SA-2A-puro-pA donor and pAAVS-CAGGS-EGFP, result in targeted integration of a 1 kb- puromycin cassette and 4.2 kb cassette (comprised of puromycin cDNA, BGH poly A, CAGGS promoter, EGFP cDNA and poly A), respectively. pSA-2A-Puro-Hybrid FVIII donor vector, assembled in our lab by inserting the hybrid FVIII cassette into the pAAVS1 SA-2A-puro-pA donor vector results in targeted integration of a 8.85 kb donor sequence (comprised of puromycin cDNA, BGH poly A, human ferritin light chain promoter, hybrid FVIII cDNA and BGH poly A).

4.4.1. Optimization in K562 cells

4.4.1.1. Gene transfer and selection

K562 cells were electroporated with 10 μg of donor plasmid DNA (pZDonor-AAVS1, pZDonor GFP or pZDonor hybrid FVIII) and either 5 μl of AAVS1 ZFN mRNA (Sigma-Aldrich) or 5 μg of AAVS1 dual ZFN plasmid DNA constructs in 100 μl of Amaxa®Cell Line Nucleofector®Kit V solution (Lonza) using an Amaxa® Nucleofector™ I device at setting T-016. An additional 2 μg of EGFP plasmid DNA was used to gauge transfection efficiencies in experiments where pZDonor-AAVS1 or pZDonor hybrid FVIII plasmids were used.

K562 cells with stable transgene expression were typically selected by culturing in media supplemented with 0.8 mg/ml of Geneticin® for 14 days before withdrawal of the selective drug.

4.4.1.2. Assessing gene transfer efficiency

4.4.1.2.1. Flow cytometry

Gene transfer efficiencies were typically evaluated 24 hours post-electroporation by fluorescence-activated cell sorting analysis using a BD FACSCalibur™ flow cytometer.

Generally, cells were spun down at 500xg for 5 min, resuspended in 600ul of PBS, filtered through 40 micron nylon filter mesh into FACS tubes and analyzed for GFP expression using 488 nm argon laser and 530/30 bandpass filter. FlowJo v7.22 software was used to estimate the total percentage of GFP positive cells.

4.4.1.3. Integration junction PCR

Integration junction PCR was performed on 200 ng of genomic DNA using DyNAzyme EXT DNA polymerase and primers specific to the integrated vector and adjacent genomic locus at the integration site. The plasmids electroporated and primers used to evaluate respective integration junctions are listed in the **Table 4** below.
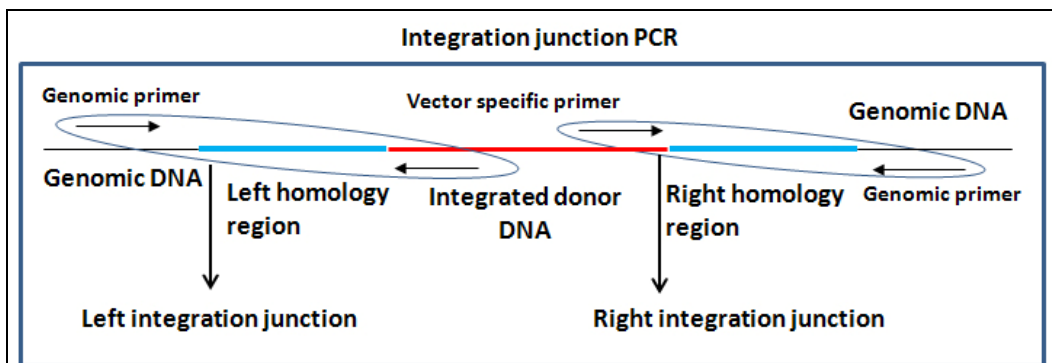
Figure 4.1   Schematic of integration junction PCR. Left and right integration junctions were amplified using a primer specific to genomic DNA adjacent to the integration site and a primer specific to the integrated vector.

Control PCR was performed with AAVS1 genomic primers (forward: 5' aagaagcgcaccacctccaggttct 3'; reverse: 5' atgacctcatgctcttggccctcgta 3'). All PCR was set up in a 20 µl reaction volume and amplified for 30 cycles using a PTC-200 Peltier gradient thermal cycler (MJ Research Inc., USA) at an annealing temperature of 62°C and extension time of 1 min per cycle. Amplified products were electrophoresed on 1% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and quantified using QuantityOne software. Integration junction PCR products were sequenced with appropriate primers to verify their identity.

Table 4   Primer sequences used for amplifying AAVS1-specific genomic integrations of different donor DNAs

Plasmid electroporated (insert/ junction being detected)	Forward primer (5' → 3')	Reverse primer (5' → 3')	Size of PCR amplicons (bp)
pZDonor (50 bp MCS)	agcttgaattctctagaaatatt ctcgaggtttaaacgtcgacg c	ggaacggggctcagtctg	1026
pZDonor-EGFP (Left junction)	ggccctggccattgtcactt 3'	cgtcaataggggggcgtacttg gcatatgatac	1237
pZDonor-EGFP (Right junction)	gacatagcgttggctacccgt gatattgctgaagagc	ggaacggggctcagtctg	1356
pZDonor-hybrid F8 (Left junction)	ggccctggccattgtcactt	cgtcaataggggggcgtacttg gcatatgatac	1237
pZDonor-hybrid F8 (Right junction)	gacatagcgttggctacccgt gatattgctgaagagc	ggaacggggctcagtctg	1356
pAAVS1-SA-2A-puro-pA donor (Left junction)	ggccctggccattgtcactt	cggtcatctcgagcctaggg	1103
pAAVS1-SA-2A-puro-pA donor (Right junction)	tcaccgagctgcaagaact	ggaacggggctcagtctg	1621
pAAVS1-SA-2A-puro-pA donor (Long PCR)	ggccctggccattgtcactt	cttggccacgtaacctgaga	3084
pAAVS-CAGGS-EGFP (Left junction)	ctggccattgtcactttgcg	cggtcatctcgagcctaggg	1099
pAAVS-CAGGS-EGFP (Right junction)	ctactcccagtcatagctgtc	cttggccacgtaacctgaga	1370
pAAVS-CAGGS-EGFP (Long PCR Left)	ctggccattgtcactttgcg	gatggggagagtgaagcaga acg	2382
pAAVS-CAGGS-EGFP (Long PCR Right)	aaacggccaccagttcagcg	cttggccacgtaacctgaga	2517

4.4.1.4. Restriction fragment length polymorphism

RFLP assay is used to estimate the frequency of site-specific integration of a 50 bp donor fragment (present in pZDonor-AAVS1) into the AAVS1 site. It gives an estimate of the percentage of the population that had successfully attained site-specific genome modifications.
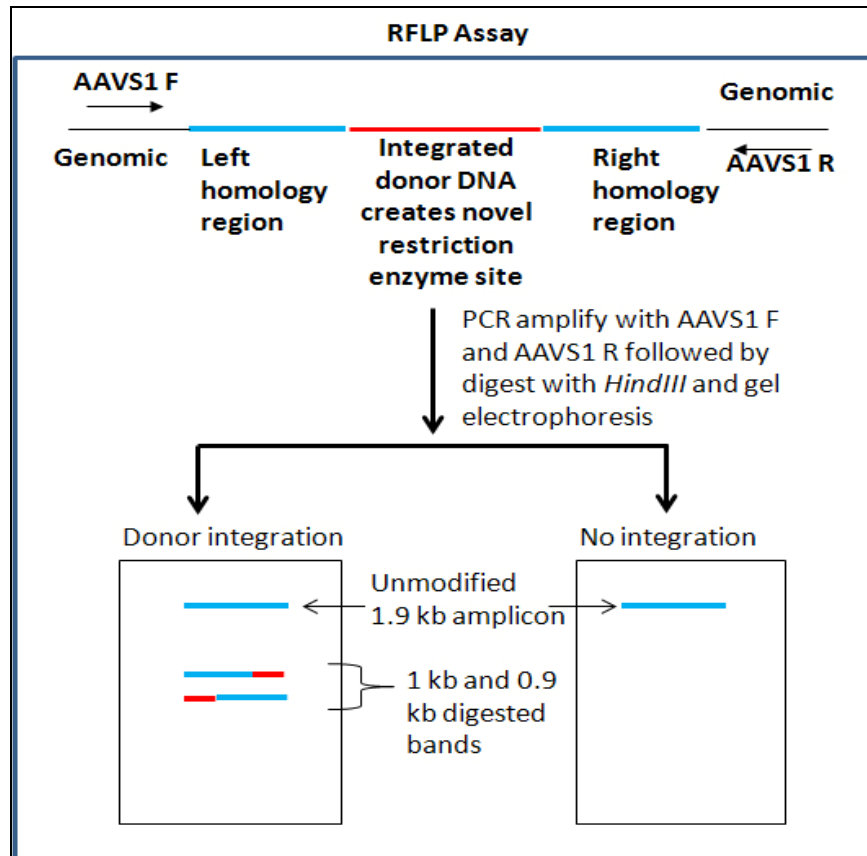
Figure 4.2   Schematic of RFLP to identify AAVS1 locus modified by site-specific integration of donor DNA. ZFN-mediated integration of 50 bp donor DNA from pZDonor results in novel restriction enzyme sites *(Hin*dIII in this example) in the PCR amplicons. Genome modification is identified by restriction enzyme digest and the appearance of smaller cleaved bands on gel electrophoresis.

Basically, approximately 200 ng of genomic DNA extracted from cells 4 days post-treatment with 10 μg of pZDonor-AAVS1 in the absence or presence of ZFNs was amplified with a pair of genomic primers spanning the AAVS1 integration site (forward: 5' ggccctggccattgtcactt 3'; reverse: 5' ggaacggggctcagtctg 3') using DyNAzyme EXT DNA polymerase. Novel restriction enzyme sites present in the 50 bp fragment of the pZDonor-AAVS1 are incorporated into the AAVS1 genomic site in cells which attain site-specific integration and will be present in the PCR amplicons if a cell population includes a subset of cells having the desired integration. PCR amplicons were digested with 1 U of *Hin*dIII at 37°C for 1 hr, resolved in 8%

polyacrylamide gels, stained with 10 µg/ml ethidium bromide and visualized with a BioRad®Gel Doc 2000 transilluminator. Unmodified genomic DNA was identified by a 1.9 kb amplicon, while site-specific genome integration was identified by the presence of 1 kb and 0.9 kb fragments. Band intensities were determined by densitometry (QuantityOne software).

4.4.1.5. Densitometric measurements

Electrophoresed junctional PCR products and RFLP products imaged with BioRad®Gel Doc 2000 transilluminator were quantified by densitometry, when necessary. The intensity and volume of DNA bands highlighted with the band volume contour tools were calculated using the QuantityOne software.

For junctional PCR, the intensity ratio of site-specific amplicon: control genomic amplicon expressed as a percentage was used as a quantitative estimate of integration efficiency.

For RFLP assay, the intensity of bands corresponding to genome modification (1 kb plus 0.9 kb fragments) expressed as a percentage of unmodified genome amplicon (1.9 kb) was used as a quantitative estimate of integration efficiency.

4.4.1.6. Direct PCR

Single cells from a bulk population of either unselected or G418/Gancyclovir selected cells were sorted by FACs into individual wells of 96-well plates and allowed to expand in culture.

Cells were lysed by incubating at 98°C for 10 min in 60 µl of cell lysis buffer with DNA release additive (1 part DNA release additive: 40 parts dilution buffer). Integration junction PCR and amplification of a control locus were performed on 2 µl of cell lysates using the Phusion Human Specimen Direct PCR kit and the following cycling parameters: 98°C for 5 min, 35 cycles of 98°C for 10s, 62°C for 10s and 72°C for 1 min, followed by a final extension at 72°C for 5 min. Primers used in the direct PCR assay were similar to those listed in section 4.4.1.3, depending on the cell population that was tested (with pZDonor EGFP or pZDonor hybrid FVIII integration). Amplified products were resolved by electrophoresis in 1% agarose gels (with 10 µg/ml of ethidium bromide) and imaged using a BioRad®Gel Doc 2000 transilluminator.

4.4.1.7. Evaluation of ZFN construct variants and mild hypothermia

Three different ZFN variants, OH (obligate heterodimers), Sharkey (OH modified according to Guo J. *et al* [9]) or Enhanced Sharkey (Sharkey ZFN variants further modified according to Doyon Y. *et al* [10]) were tested to compare their ability to induce site-specific genomic cleavage and promote homologous recombination of a 50 bp donor DNA provided from pZDonor-AAVS1 plasmid DNA. K562 cells were co-electroporated with 2 μg of pEGFP-C1 and 10 μg of pZDonor-AAVS1 only or with 5μg of one of the following AAVS1 ZFN variants; OH (obligate heterodimers), Sharkey or Enhanced Sharkey ZFN and cultured at either 37°C for 4 days or at 37°C for 1 day followed by 30°C for 3 days. Junctional PCR and RFLP assays were performed to determine homologous recombination of donor DNA at the AAVS1 locus. Densitometric measurements of RFLP products were used to estimate the proportion of cells that attained targeted gene integration.

4.4.1.8. Evaluation of donor insert size and ZFN dose

The ability to integrate larger donor DNA fragments at the AAVS1 locus was investigated by co-electroporating K562 cells with pZDonor EGFP (3 kb donor fragment) or pZDonor hybrid FVIII (9 kb donor fragment) and Enhanced Sharkey AAVS1 ZFN plasmid DNA. Cells electroporated with donor DNA only served as negative controls. Electroporated cells were either unselected or selected with 1mg/ml G418 for 7 days before being expanded for 1 month. Genomic DNAs extracted from treated cells were templates for integration junction PCR (left and right junctions) and control PCR.

4.4.2. Evaluation of ZFN modification of primary cells

The ability to induce AAVS1 site-specific genomic cleavage and promote homologous recombination-mediated integration of donor DNA was investigated in primary human CLECs, bone marrow-derived stromal cells and adipose-derived stromal cells by co-electroporating 10 μg of pZDonor-AAVS1 and 7 μg of Enhanced Sharkey AAVS1 ZFN plasmids using optimized electroporation settings. Junctional PCR was used to determine site-specific integration. Where there was evidence of homologous recombination, RFLP assay was performed to estimate the proportion of cells attaining site-specific integration of donor DNA.

4.4.2.1. Primary cells and culture conditions

Culture conditions for CLECs, BMSCs, ADSCs, fibroblast and HSCs are described in section 4.1.4.

4.4.2.2. Gene transfer

Gene transfer was performed by electroporation, initially using an Amaxa® Nucleofector™ I device and later using an Amaxa® 4D Nucleofector™ device and their respective cuvettes. Generally, 2 to 10 million cells were electroporated with 2 to 10 µg of plasmid DNA in 100 µl of the indicated Amaxa solutions.

CLECs were nucleofected using either Amaxa™ Basic primary mammalian epithelial cell solution and one of the following settings; S-05, T-23, U-17, T-13 or T-20 using the Nucleofector™ I device or in nucleofector Primary cell solution P1 and with one of the following settings; CM102, CM113, EA104, ED100 or DS109 using the Nucleofector™ 4D device. The optimal setting for Nucleofector™ I and Nucleofector™ 4D devices were determined to be T-23 and CM113, respectively.

Four different lines of primary human bone marrow stromal cells (Lonza MSC, BMSC1, BMSC2, Gan BMSC) and adipose-derived stromal cells (ADSC1, ADSC2) were electroporated using Amaxa™ Basic primary mammalian bone marrow cell solution and pulses delivered by the indicated program settings (C16, C17, U-23) with a Nucleofector™ I device.

Three different lines of primary human fibroblasts, Hs68, KF1 and NF123, were electroporated using Amaxa™ Basic primary mammalian fibroblast cell solution and pulses delivered by the indicated program settings (A-24, U-23, T-16, V-13) with a Nucleofector™ I device.

4.4.2.3. RT-PCR evaluation of *PPP1R12C* transcript expression

Expression of *PPP1R12C* transcripts in K562 cells, primary human fibroblasts, a human embryonic stem cell line (ES) (cells provided by Dr. Mark Richards, Nanyang Polytechnic, Singapore) and CLECs were evaluated by RT-PCR. RNA extracted from cells (RNeasy mini kit) was treated with DNAse I, reverse transcribed (SuperScript™ II reverse transcriptase) and amplified using DynazymeEXT DNA polymerase and the following PCR primers (see **Appendix 1**): *PPP1R12C* exons 4 – 6 (forward primer 5′ agaggaattgctccttcatgacac 3′; reverse primer 5′ caacaggctcagtacttcctcatc 3′); *PPP1R12C* exons 8 - 12 (forward primer 5′ gagctctgtgtgtcgtctg 3′; reverse primer 5′ ccgtggaggctgtgggga 3′); and γ-actin (forward primer 5′ accactggcattgtcatggactct 3′; reverse primer 5′ atcttgatcttcatggtgctgggc 3′).

Amplified products were electrophoresed on 2% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and quantified using QuantityOne software.

4.4.2.4. Evaluation of AAVS1 ZFN mRNA and protein expressions

CLECs electroporated with plasmid DNA encoding both left and right AAVS1 ZFNs were investigated for expression of ZFN mRNA transcripts and protein by RT-PCR and immunoblotting, respectively.

A time course study (8 – 144 hr post-electroporation) was performed to determine the temporal profile of ZFN mRNA. Total RNA extracted at the indicated time points were DNAse treated, reverse-transcribed and PCR amplified using primers indicated in **Appendix 1**). Samples which had not been reverse transcribed (minus RT) were PCR amplified as above. Amplified products were electrophoresed on 1% agarose gels, imaged using BioRad®Gel Doc 2000 transilluminator and quantified using QuantityOne software. Densitometric measurements of ZFN transcript bands were normalized to the respective actin levels and expressed as a percentage of ZFN mRNA levels observed at 8 hours.

For evaluation of ZFN proteins, transfected CLECs were incubated either at 37°C for 3 days, 30°C for 3 days or 37°C for 1 day followed by 30°C for 2 days, prior to protein extraction. Detection of FLAG-tagged ZFN protein and β-actin protein (as loading control) was by immunoblotting with 1:1000 diluted monoclonal ANTI-FLAG® M2 antibodies and 1:2000 diluted monoclonal anti-β-actin antibody, respectively, as detailed in section 4.5.5.

4.4.2.5. Evaluating site-specific genomic cleavage using *Cel*-1 nuclease

In the absence of donor DNA, ZFN-mediated site-specific cleavage of genomic DNA is usually repaired by the error-prone NHEJ pathway that results in indel formation at sites of genomic cleavage and repair. The capacity for NHEJ mutagenic repair was evaluated by performing a *Cel*-1 nuclease assay using the Surveyor™ mutation detection kit (Transgenomic Inc., USA) according to the manufacturer's protocol. The *Cel*-I mismatch nuclease only cleaves when a mismatch is present in amplicons [21] and has a reported sensitivity of detecting mutants as infrequent as 1 in 10000 copies. AAVS1 genomic region to be evaluated was amplified using the following primer pair: Cel-1 forward 5' ttcgggtcacctctcactcc 3'; Cel-1 reverse 5' ggctccatcgtaagcaaacc 3'. Approximately 200 ng of gel-purified PCR amplicon was denatured at 95°C for 5 min and reannealed by cooling to room temperature followed by *Cel*-1 digest at 42°C for 40 min. *Cel*-1 digested reactions were resolved in 10% polyacrylamide gel, post-stained with 10 µg/ml ethidium

bromide and visualized with a BioRad®Gel Doc 2000 transilluminator. Band intensities were determined by densitometry (QuantityOne software). Unmodified genomic DNA was identified by a 469 bp amplicon while site-specific genome cleavage and repair was identified by the presence of 287 bp and 182 bp fragments. The proportion of cleaved bands to uncleaved PCR products gave an estimate of the proportion of mutant cells that arose from NHEJ repair, and thus, by inference, the efficiency of genomic cleavage.
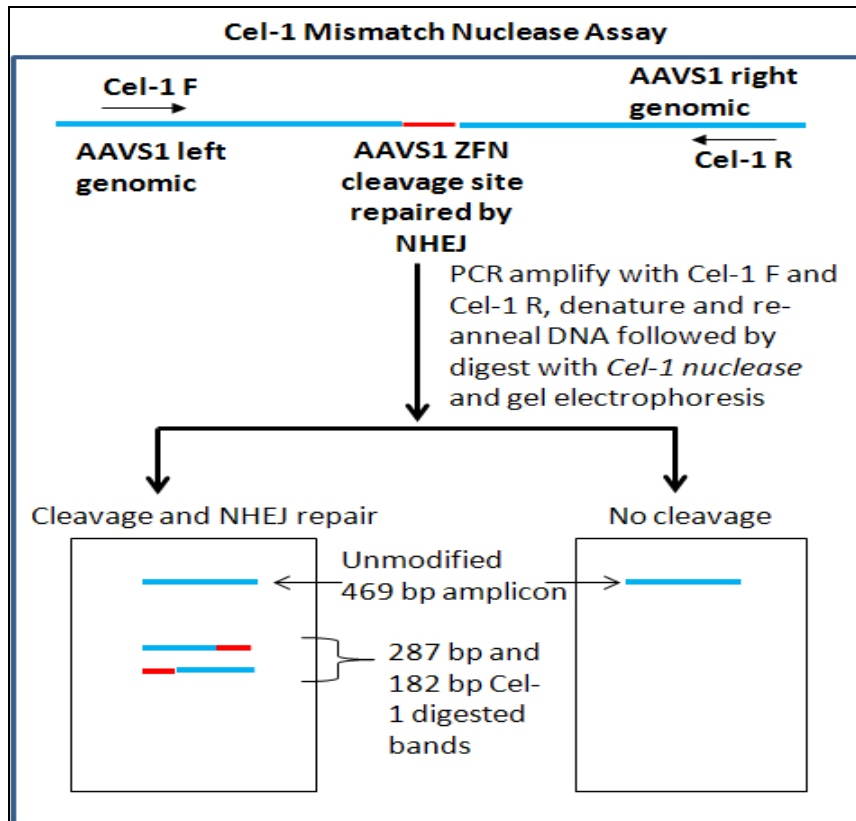
Figure 4.3 Schematic of *Cel*-1 mismatch nuclease assay. Site-specific double-stranded DNA breaks are repaired by error prone non-homologous end joining (NHEJ) resulting in insertions and/or deletions (indels) at sites of repair. The genomic locus to be evaluated for evidence of genomic cleavage and repair is amplified using a pair of genome specific primers (Cel-1 F and Cel-R). Amplicon DNA is denatured and annealed to allow heteroduplex formation, digested with *Cel*-1 mismatch nuclease and resolved polyacrylamide gel electrophoresis. In this example, genomic DNA cleavage and subsequent NHEJ repair is identified by cleaved fragments (287 bp and 182 bp) that result from the presence of a mismatch in the unmodified (i.e. uncleaved) 469 bp PCR amplicon.

4.4.2.6. Evaluating site-specific integration

Site-specific integration of pZDonor-AAVS1 (50 bp donor), pZDonor EGFP (4 kb donor), pZDonor hybrid FVIII and pZDonor hybrid FVIII TK007 (both 9 kb donors), AAVS1-SA-2A-Puro-pA (1kb donor) , AAVS-CAGGS-EGFP (4.2 kb

donor) and pSA-2A-Puro-Hybrid FVIII (8.85 kb donor) were evaluated following co-transfection with a bicistronic AAVS1 ZFN variant (Enhanced Sharkey) plasmid DNA. CLECs with stable integration of pZDonor GFP, pZDonor hybrid FVIII and pZDonor hybrid FVIII TK007 resulted in stable expression of the neomycin resistance gene and were selected 4 days post-electroporation with 1 mg/ml G418 for 7 days. Similarly, CLECS with integration of AAVS1-SA-2A-Puro-pA , AAVS-CAGGS-EGFP and pSA-2A-Puro-Hybrid FVIII which were puromycin-resistant were selected 4 days post-electroporation with 0.5 μg/ml puromycin for 7 days. Cells integrated with the pZDonor hybrid FVIII TK007 were further selected by culturing in medium containing 1 μM gancyclovir for 7 days. Cells with random integration of the donor vector expressed TK007 and therefore will be killed by gancyclovir, while cells with site-specific integration of donor DNA by homologous recombination do not retain the TK007 cassette and are resistant to gancyclovir selection. Genomic DNA was extracted from approximately 2 million stable cells using either FavorPrep™ Blood genomic DNA extraction mini kit (Favorgen Biotech Corp., Taiwan) or QIAamp® DNA Blood mini kit. Site-specific integration of the various plasmid vectors were evaluated by integration junction PCR and/or RFLP assay as described in sections 4.4.1.3 and 4.4.1.4.

4.4.2.7. Efficiency and accuracy of integration

The efficiency or frequency of site-specific integration of pZDonor-AAVS1 was evaluated by densitometric measurements of RFLP assay products as described in section 4.4.1.4. Efficiency of site-specific integration of pZDonor hybrid FVIII or pZDonor hybrid FVIII TK007 was evaluated by direct integration junction PCR on FACS sorted clonal cells. In brief, either transiently or stably transfected cells derived by G418 or gancyclovir selection were FACS sorted (4 cells per well) into 96-well flat bottom tissue culture plates and allowed to expand for 10 days. Direct PCR to amplify the right integration junction and control locus were performed on the sorted cells as detailed in section 4.4.1.6. The proportion of clonal cells positive for integration junction compared to those positive for control locus amplification was the estimated frequency of cells attaining site-specific integration. The effect of G418 and gancyclovir selections in increasing the frequency of cells with site-specific integrations was tested using this method.

Integration junction PCR products were sequenced with vector and genomic DNA specific primers to confirm their identity and to determine if any deletions or insertions had occurred at the integration junctions.

4.4.2.8. Durability of FVIII secretion *in vitro*

Cells (bulk population or clonal) stably modified by integrating hybrid FVIII cDNA and identified as positive for site-specific integration by both left and right integration junction PCR were monitored for durable FVIII secretion *in vitro*. Approximately 100000 cells were seeded onto a single well of a 12-well tissue culture plate (done in triplicate) and cultured in 500 µl of medium overnight. Cell culture supernatant was harvested 24 hours later and stored at -80°C until time of assay. Re-seeding of the same cells and collection of conditioned media were repeated weekly for the 1st month followed by once per fortnight for the following month. FVIII activity in conditioned media was quantified using the Coamatic FVIII assay described in section 4.3.3.1. Naive CLECs, CLECs transfected with episomal plasmid only and CLECs with random integration of hybrid FVIII were also monitored for durability of FVIII secretion.

4.4.2.9. RT-PCR and genomic PCR evaluation for HSV-TK and HSV-TK.007 transcript and genomic integration

Expression of HSV-TK and HSV-TK.007 transcripts in transfected CLECs were evaluated by RT-PCR. RNA extracted from cells (RNeasy mini kit) was treated with DNAse I, reverse transcribed (SuperScript™ II reverse transcriptase) and amplified using DyNAzyme EXT DNA polymerase and PCR primers indicated in **Appendix 1**. Samples which were not reverse transcribed (minus RT) were PCR amplified as above. Genomic integrations of HSV-TK and HSV-TK.007 were evaluated by PCR of 200 ng of genomic DNA using the same primers above. Amplified products were electrophoresed on 2% agarose gels and imaged using BioRad®Gel Doc 2000 transilluminator and quantified using QuantityOne software.

4.4.3.   Biosafety evaluation of ZFN modified cells

ZFN modified cells (bulk or clonal culture) identified as positive for site-specific integration of hybrid FVIII cDNA were selected for evaluation of potential genotoxicity acquired during the course of genome modification and cell expansion *in vitro*.

4.4.3.1. Immunofluorescence staining for histone H2AX

Phosphorylated histone H2AX (S209) is an established marker for double-stranded DNA breaks (DSDB) and serves as an indicator of genotoxicity. Electroporated cells were assessed for DSDBs by immunostaining for phosphorylated

H2AX 2 days post-treatment. Cells were trypsinized, fixed with 3.7% formaldehyde in PBS for 10 min and 90% methanol at -20°C for 2 hrs, permeabilized (0.5% Triton-100, 2% BSA in PBS) for 10 min and incubated with 1:40 diluted (in 2% BSA) anti-phosphohistone H2AX (Ser139) (20E3) rabbit mAb (Alexa Fluor® 647 conjugated) (Cell Signalling Technology®, USA) for 1 hour at 25°C. Cells were washed twice with PBS, resuspended in 500 µl of PBS, filtered through 40 micron nylon filter mesh into FACS tubes and analyzed using 633 nm He-Ne laser for excitation and 661/16 nm bandpass filter for detection. FlowJo v7.22 software was used to determine the percentage of phosphoH2AX-positive cells. Untreated cells served as negative controls while cells treated with 10 µM etoposide for 1 hour served as positive controls for DSDB.

4.4.3.2. Viability assay, MTS assays and *in vitro* colony formation assay

Cell viability was assessed either by trypan blue exclusion manual cell counts or MTS assay 24-48 hours post-electroporation. Adherent cells were detached, washed once with PBS and resuspended in 1 ml of PBS.

For trypan blue exclusion cell counts, 10 µl of an equal mixture of cell suspension and trypan blue was loaded by capillary action into a Neubauer hemocytometer. The number of unstained cells that did not take up trypan blue (live cells) and number that stained blue (dead cells) were counted. The number of live and dead cells in each sample was determined in triplicate. Viability was expressed as the percentage of live cells to the total cell number.

For MTS assay (CellTiter 96® AQ$_{ueous}$ One Solution Cell Proliferation Assay), approximately 500-1000 cells in 100 µl of culture medium were seeded into each of quadruplicate wells of a flat bottom 96-well tissue culture plate. Naive wild type cells, cells electroporated without any DNA or with donor DNA only served as controls to compare against cells electroporated with both donor DNA and ZFN. Twenty-four hours later, 20 µl of CellTiter 96® AQ$_{ueous}$ One Solution reagent was added to each well and incubated for 1-4 hours at 37°C in a humidified 5% CO_2 incubator, after which reduction of the tetrazolium compound, MTS, to a colored formazan product was quantified by absorbance readings at 490 nm in a Dynex MRX II 96-well plate reader. Dead cells are unable to reduce MTS, while metabolically active live cells reduce it through the action of mitochondrial reductases into a purple-colored formazan product.

In vitro colony formation assay was performed as detailed in section 4.3.2.9 except that 100 cells were initially seeded into each well of a 6-well plate and allowed to grow for 14 days before crystal violet staining was performed.

4.4.3.3. RT-qPCR analysis of CLEC stables with puro FVIII transgene integration

Reverse transcription (RT)-quantitative PCR was performed as detailed in section 4.3.4.2.5 and with primers indicated in **Appendix 1** to determine changes in the levels of specific transcripts in ZFN-treated puromycin-resistant CLECs with transgene integration at the AAVS1 locus (derived from electroporating pAAVS1 SA-2A-puro-pA donor, pAAVS-CAGGS-EGFP or pSA-2A-Puro-Hybrid FVIII). CLECs electroporated without plasmid DNA and of the same number of population doublings served as controls. Intron-spanning exonic primers were used to amplify the endogenous *PPP1R12C* transcript (exons 4-6), neighboring genes within 1-Mb of AAVS1 integration site (*LILRB4*, *ISOC2*, *PPP6R1*, *NAT14*, *ZNF579*, *FIZ1* and *RDH13*), potential interacting partners of *PPP1R12C* predicted by Gene Network Central™ and Spring 9.05 which were identified by transcriptome data as significantly altered (*DUSP1*, *DUSP6*, *CDC6* and *DUSP16*) and a housekeeping gene, glyceraldehydes-3-phosphate dehydrogenase (*GAPDH*). Transcript levels were normalized to *GAPDH* expression levels and the fold-change in transcript levels in CLECs with transgene integration was reported relative to transcript levels in control CLECs, using the 'delta-delta $C_{(T)}$ method'[18].

4.4.3.4. Transcriptome analysis of stable CLECs with targeted integration of puro FVIII cassette

Transcriptomes of naïve unmodified CLECs and bulk population of ZFN-treated puromycin-resistant CLECs with targeted integration of hybrid FVIII cassette were determined using HU133 plus 2.0 array and analyzed using GeneChip Operating Software (Affymetrix). Transcripts whose expression levels differed significantly (determined by Wilcoxon signed-rank test) by more than two-fold in genome-modified CLECs compared to naïve CLECs were considered significantly altered and were further analyzed. DAVID 2.1 Functional Annotation Tool (http://david.abcc.ncifcrf.gov) was used to ascribe functions and other annotations for significantly altered transcripts and for pathway mapping. Significantly altered transcripts were also referenced to a compilation of known proto-oncogenes and tumor suppressor genes (http://www.bushmanlab.org/links/genelists) in order to ascertain if they belonged to either category. Potential interacting partners of PPP1R12C were predicted using Gene Network Central™ (http://www.sabiosciences.com/genenetwork/genenetworkcentral.php) and String 9.0 (http://string-db.org/).

4.4.3.5. Deep sequencing to evaluate top-10 potential off-target sites for ZFN activity

Previously identified and published 10 most likely potential off-target sites for AAVS1 ZFNs[3] (**Appendix 10**) were evaluated by massively parallel deep sequencing using Illumina MiSeq sequencing platform (150 bp paired-end sequencing). Amplicons for AAVS1 ZFN-binding site (AAVS1 locus) and the top-10 potential off-target sites (OT1 – OT10) were amplified from genomic DNA extracted from naïve unmodified CLECs and ZFN-treated puromycin-resistant CLECs with targeted integration of hybrid FVIII cDNA at the AAVS1 locus using Dynazyme™ EXT DNA polymerase and primers listed in **Appendix 1**. Amplicons were resolved by agarose gel electrophoresis and purified using QIAquick Gel Extraction kit (Qiagen). A commercially synthesized synthetic DNA fragment (GenScript, USA) similar to the AAVS1 locus sequence except for a 5 bp deletion between the ZFN binding site was spiked into wild type AAVS1 locus amplicon at the following mass ratios of 1:10, 1:100, 1:500 and 1:1000 to determine the sensitivity of the method to detect indels. The amplicons we generated were outsourced to AITbiotech Pte. Ltd., Singapore for library construction (Nextera® XT DNA Sample Preparation Kit, Illumina), sequencing and bioinformatic analysis.

Paired-end sequencing reads were aligned to the reference human genome assembly (version hg19) using Burrows-Wheeler Aligner[22] (BWA). Indels were identified using SAMtools (http://samtools.sourceforge.net/). Indels that occurred within as well as 10 bp upstream or downstream of the SELEX predicted sequence (**Appendix 10**) were evaluated as potential off-target events. Normalized frequency of indels (indel frequency divided by total number of aligned reads at the position) determined from the spike-in experiment ranged between 0.0565 to 4.13. Thereafter, only indels with normalized frequency of over 0.05 were considered significant for both untreated CLECs and puro FVIII ZFN stable CLECs.

4.5.   Molecular biology techniques

4.5.1.   Plasmid DNA isolation

All plasmid DNA used in this study were prepared using Qiagen plasmid DNA isolation kits, according to the manufacturer's protocol. Typically, cloned or commercially purchased plasmids were transformed by the heat-shock method into competent DH5α *E. coli* and plated on LB agar plates containing appropriate selective antibiotics. Starter cultures from a single bacterial colony, cultured for a minimum of 8 hr at 37°C, were diluted 1:500 into LB media (usually 200 ml for a maxi

preparation) containing appropriate antibiotics and cultured overnight with shaking at 37°C. Plasmid DNA harvested from bacterial cultures were dissolved in Ultrapure™ DNase-RNase free distilled water and stored at -20°C.

4.5.2. RNA isolation

Total RNA was isolated from cultured cells using either TriPure isolation reagent (followed by phenol-chloroform clean-up and sodium-acetate/ethanol precipitation) or using RNeasy® Mini kit (Qiagen) according to the supplier's protocol. When necessary, total RNA (1 μg) was digested with 1U of RQ1 RNase-free DNase at 37°C for 1 hr to eliminate co-purified DNA. DNase-treated RNA was then purified using RNeasy® Mini kit, eluted with Ultrapure™ DNase-RNase free distilled water and stored at -80°C.

4.5.3. Genomic DNA isolation

High molecular weight genomic DNA was isolated from cultured cells using Qiagen Genomic-tip 100/G (up to 100 μg of DNA) or QIAamp DNA mini kit, according to the manufacturer's protocol, eluted in Ultrapure™ DNase- and RNase-free distilled water and stored at -20°C.

4.5.4. Polymerase chain reaction (PCR)

PCR was routinely used in this study for DNA cloning and diagnostic amplification/screening of 1st strand cDNA or genomic DNA. Generally, PCR was performed using DyNAzyme EXT DNA polymerase on 10 ng of plasmid DNA, 200 ng of genomic DNA or 2 μl of 1st strand cDNA products in a total volume of 20 μl. A typical PCR setup consisted of 1x PCR reaction buffer with 1.5 mM $MgCl_2$, 1 μM each of forward and reverse primers, 2 mM of each dNTP and 1U of DNA polymerase. PCR was performed using a PTC-200 Peltier gradient thermal cycler (MJ Research Inc., USA).

4.5.5. Reverse transcription reaction

Reverse transcription of RNA for 1st strand cDNA synthesis was generally performed on 500-1000 ng of total RNA using either oligo(dT)$_{12-18}$ or random primers and SuperScript™ II reverse transcriptase or using iScript™ Advanced cDNA synthesis kit, according to the manufacturer's protocol.

4.5.6. Protein immunoblotting

Protein samples for analyses were extracted from 2 to 10 million cells by cell lysis using M-PER mammalian protein extraction reagent (Pierce, USA) and quantified using a BCA protein assay kit. Twenty to 50 μg of protein from each cell lysate was mixed 4:1 with 5x bromophenol loading dye (60 mM Tris HCL at pH6.8, 20% glycerol, 2% SDS, 0.01% bromophenol blue and 5% β-mercaptoethanol), denatured at 95°C for 10 min, resolved at 90V for 1 hour on 14% SDS-PAGE under reducing conditions using a BioRad Protean II vertical electrophoresis system. Proteins resolved by electrophoresis were electrotransferred at 4°C onto nitrocellulose membranes (Bio-Rad Laboratories) at 90 V for 1 hour using a BioRad Trans-blotting system. Membranes were subsequently blocked with 5% non-fat milk in 0.1% (v/v) Tween 20 in PBS for 1 hour, rinsed with wash buffers (Tris-buffered saline, pH7.6, with 0.1% (v/v) Tween 20) and probed with specific primary antibodies diluted in PBS with 1% non-fat milk and 0.1% (v/v) Tween20 for 1 hour at room temperature. Membranes were washed four times with wash buffer before and after a 1 hour incubation with diluted horseradish peroxidase-conjugated goat anti-mouse or goat anti-rabbit secondary antibodies (Promega and Santa Cruz Biotechnology, respectively), incubated with a chemiluminescence-based photoblot substrate reagent (Amersham Biosciences, Piscataway, NJ) and developed in a dark room by exposing to autoradiography film (Hyperfilm ECL, Amersham) for 1 to 30 min before developing with a Kodak film processor.

4.5.7. DNA sequencing

DNA sequencing was performed using BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystem, Life Technologies Corp) with 3.2 pmol of appropriate primer and either 50-100 ng of purified PCR product or 150-300 ng of double-stranded plasmid DNA as the template. Briefly, each 20 μl sequencing reaction was reacted for 30 cycles of 95°C for 10 sec, 50°C for 10 sec and 60°C for 4 min. Reactions were terminated with 0.1 volume of 3M sodium acetate and 0.1 M EDTA, precipitated with 2 volumes of absolute ethanol, washed with 70% ethanol and vacuum dried. Sequencing reaction products were outsourced to 1st BASE Pte Ltd (Singapore) for electrophoresis and sequencing results were analyzed using BioEdit software.

4.5.8. Gel electrophoresis

DNA samples mixed with 6x DNA loading dye to a final working concentration of 1x (total volume <20 μl) were loaded onto 1% agarose gels with 10 μg/ml of ethidium bromide and electrophoresed using a Subsystem 70 electrophoresis device (Labnet International, Inc., USA) at 90 V with power supplied from a BioRad power pack for approximately 30 min using Tris-acetate-EDTA buffer. The following DNA markers were used as appropriate: 1kb DNA ladder, 100 bp DNA ladder and 50 bp DNA ladder. Electrophoresed samples were visualized and imaged using a BioRad®Gel Doc 2000 transilluminator.

RFLP assay and *Cel*-1 nuclease digest products were resolved in 5% polyacrylamide gels (30% acrylamide mix 29:1, 0.1% ammonium persulfate, 0.05% TEMED, 1xTBE) using Tris-borate-EDTA buffer in BioRad Protean II vertical electrophoresis cells system at 90V for approximately 1 hr, stained with 10 μg/ml ethidium bromide or 1x SyBr Gold for 30 min and imaged using a BioRad®Gel Doc 2000 transilluminator.

4.6. Cell biology techniques

4.6.1. Microscopy

4.6.1.1. Light and fluorescence microscopy

A brightfield/phase contrast microscope (CK30-F200, Olympus) was used for routine experiments. Phase contrast mode was used to enumerate cells using a hemocytometer while bright field mode was used to view stained tissue sections, cytospun cells and G-banded metaphase chromosomal preparations. Imaging of stained tissue sections and cells was done with an Axiovert 25CFC microscope and KS400 software (Carl Zeiss).

Inverted bright-field microscope (Nikon Eclipse TS100) was routinely used to monitor cell cultures for confluency and visual signs of microbial contamination. Bright-field images of cells were obtained using NIS-Elements 3.0 software.

EGFP fluorescent cells were visualized and imaged using an inverted Nikon Eclipse TE-300 fluorescence microscope and Nikon ACT-1 software. SKY paint-labeled chromosomes and cells hybridized with fluorescent probes for FISH studies were viewed and imaged using an Olympus BX61 epifluorescence microscope and analyzed using BandView® software.

Unless otherwise stated, all general observations and viewing were done using either a 10x or 40x objective lens magnification. Where specified, imaging was done with 60x or 100x objective lens magnification.

4.6.2. Histology

4.6.2.1. Tissue processing

Tissues were rinsed with PBS and fixed in 10% formalin (in buffered PBS) at room temperature for at least 6 hours before mounting in tissue processing cassettes (Thermo Scientific) and processing in an automated Leica tissue processor. A typical tissue processing procedure for small tissues would include 5 min sequential dehydration in 70%, 80% and 90% ethanol followed by 4 sequential 10 min incubations in absolute ethanol and 3 sequential 10 min incubations in xylene, terminating in an overnight incubation in 60°C paraffin wax. Processed tissues were transferred to 60°C paraffin wax in a Leica tissue embedding station (Leica Microsystems), embedded in a tissue cassette with paraffin wax and cooled on a 4°C cooling platform for at least 15 min until the paraffin wax solidified. Tissue embedded paraffin blocks were stored at 4°C until tissue sectioning.

4.6.2.2. Paraffin sectioning

Tissue embedded paraffin blocks were cut into 5 or 10 micron sections using a microtome blade and Leica rotary microtome equipment (Leica RM2235). Paraffin sections were straightened by placing them on a microscope slide containing a small volume of absolute ethanol and allowed to float immediately on a 42°C water-bath. Tissue sections were then collected onto pre-cleaned poly-L-lysine coated microscope slides and baked overnight at 50°C before immunohistochemical staining.

4.6.2.3. Preparation of cells by Cytospin protocol

Cells required for immunohistochemical staining were fixed in 10% formalin for 10 min, rinsed and resuspended in PBS at a concentration of 50000 cells per ml. Approximately 500 μl of cell suspension was loaded onto a Cytospin chamber that was clamped onto a microscope slide and spun in a Shandon Cytospin centrifuge (Thermo Scientific, USA) at 1500 rpm for 15 min. Slides were stored in a humidified chamber at 4°C until use.

4.6.2.4. Immunohistochemical staining

Paraffin tissue sections on slides were de-paraffinized twice with xylene for 5 min each and hydrated by sequential 2 min incubations in 100% ethanol, 95 % ethanol and 70% ethanol. Slides were then rinsed in tap water, and where antigen retrieval was necessary, heated (by microwaving) with 0.1M citrate buffer (10 mM citric acid, 0.05% Tween 20, pH 6.0) for 20-40 min. Slides were then treated with 3%

hydrogen peroxide for 15 min to inactivate endogenous peroxidase activity, rinsed with PBS and incubated at room temperature for 1 hour with diluted primary antibody. Sections were rinsed with PBS before and after a 30 min room temperature incubation with diluted secondary antibody. Dako REAL™ antibody diluent, used to dilute both primary and secondary antibodies, obviated the need for an additional blocking step. Antibody binding was visualized using Dako REAL™ EnVision™ detection system (Peroxidase/DAB+ rabbit/mouse) according to the manufacturer's protocol. Excess DAB substrate was rinsed off with PBS and tissue sections were counterstained with Mayer's hematoxylin for 3 min followed by a rinse in running tap water and sequential 2 min incubations in 70% ethanol, 95% ethanol, 100% ethanol and xylene. Slides were mounted with DPX solution, cover-slipped and observed for staining using a brightfield microscope.

4.6.2.5. Immunofluorescence staining

For immunofluorescence staining, formalin-fixed cells were cytospun onto microscope slides, further fixed in 90% methanol at -20°C for at least 2 hours, air dried and permeabilized (0.5% TritonX100, 2% BSA, 0.02% NaN^3, PBS) at room temperature for 10 min. Cells were treated with Image-iT™ FX signal enhancer (Alexa Fluor® SFX kits, Invitrogen) for 30 min at room temperature, rinsed with PBS and blocked with blocking buffer (2% BSA, 0.2% Triton X-100 in PBS) for 30 min at room temperature. Cells were incubated with 1:50 diluted primary antibodies [(Phospho-histone H2A.X (Ser139)(20E3) rabbit mAb (Alexa Fluor®647 conjugate)] for 1 hr at 37°C, rinsed with 4 washes of 2xSSC and then incubated with 1:500 diluted secondary antibodies (goat anti-rabbit IgG-Alexa Fluor® 594). PBS containing 1%BSA and 0.1% Triton X-100 was used as diluent. Both incubations proceeded at 37C for 1 hour. Following four washes with 2xSSC, cells were mounted with VECTASHIELD® Mounting Medium with DAPI (Vector Laboratories Inc., USA) and examined with an Olympus BX61 epifluorescence microscope (excitation =590 nm, emission = 617 nm).

4.6.3. Flow cytometry

Flow cytometry using detection settings in parentheses was used in experiments where analyses of cells positive for GFP cells (FL-1), APC-BrdU (FL-2), H2AX (FL-4) and propidium iodide (FL-2) were required.

Generally, stained and unstained cells were pelleted at 500xg for 5 min, resuspended in 600 µl of PBS, filtered through 40 micron nylon filter mesh into

FACS tubes, analyzed in a FACSCalibur™ flow cytometer and FlowJo v7.22 software. Samples were analyzed in triplicate.

4.7. Animal studies

All animal studies were conducted in accordance with ethical and international animal use guidelines following approval by SingHealth Institutional Animal Care and Use Committee (IACUC).

4.7.1. Anesthesia

Anesthesia was induced by intraperitoneal injection of 0.1 ml of a mixture consisting of equal parts of Hypnorm [fluanisone (10 mg/ml) and fentanyl citrate (0.315 mg/ml); Janssen Pharmaceutica, Berchem, Belgium] and Dormicum (midazolam, 5 mg/ml; Roche, Basel, Switzerland), diluted in two parts of water. Anesthetized mice were allowed to recover from surgery and anesthesia under a heat lamp for 4-6 hours.

4.7.2. Retro-orbital venous blood sampling

Capillary tubes flushed with 0.1M sodium citrate were used to collect approximately 0.05 ml of whole blood from the retro-orbital plexus of anesthetized mice into 1.5 ml tubes containing 5 μl of 0.1 M sodium citrate. Topical thrombin was applied at the site of retro-orbital puncture to stop bleeding. Plasma samples prepared from citrated whole blood samples by centrifuging at 10 000 rpm and 4°C for 10 minutes were used immediately or transferred to a new tube, flash frozen and stored at-80°C until use.

4.7.3. Implantation of cells and excision of tissues

For sub-cutaneous implantation, the skin over the dorsal neck was shaved, cleaned with 70% ethanol and scrubbed with povidone iodine solution. Five to 20 million cells in no more than 200 μl of PBS or Matrigel™ were injected beneath the epidermal layer into the subcutaneous region using a 22-G needle and syringe.

For intramuscular implantation, anesthetized mice were shaved over a hind limb, swabbed with 70% ethanol and povidone iodine solution, and a skin incision made to expose the gastrocnemius or quadriceps muscles. Cells were injected into either the gastrocnemius or quadriceps muscle using a 27-G needle and syringe. The skin incision was closed with non-absorbable sutures and the sutured area swabbed again with povidone iodine solution. External sutures were removed after one week.

For implantation *via* intravenous injection, no more than 2 million cells suspended in 200 μl of sterile PBS was injected over 30s into the tail vein of anaesthetized mice using a 29-G needle and syringe. Mice were monitored for abnormal breathing and heart beat until recovery from anesthesia.

At the termination of experiments, tissues at implantation sites were excised from euthanized mice using forceps and scissors, collected and stored in 10% formalin until processing.

4.7.4. Tail-bleed phenotypic correction assay

Tail-bleed phenotypic correction assay was performed on mice 3 days post-treatment as follows. The tail of an anesthetized mouse was placed in physiological saline solution at 37ºC for 2 minutes and then clipped 3 mm from the tip with a sharp sterile blade. The clipped tail was immediately placed in a fresh tube containing 500 μl of physiological saline solution at 37ºC and allowed to bleed (or clot) without intervention for 15 minutes. The procedure was terminated after 15 minutes and the severed tail cauterized to stop bleeding. Blood loss was quantified by the hemoglobin content of the saline solution at the end of the 15-minute period determined by absorbance measurement at 575 nm using a spectrophotometer [20].

4.8. Statistical analyses

All statistical analyses were done using GraphPad Prism program (GraphPad Software Inc., USA). Results were usually expressed as mean ± SEM, with "n" indicating the total number of samples or measurements. Analysis of variance (ANOVA) and Tukey–Kramer test were used to compare the means of three or more groups. Student's unpaired *t*- test with two-tailed *P*-values and 95% confidence interval was used for comparison between two groups (assumed to have equal variance), with Mann-Whitney test used when variances were not assumed to be equal. Fisher's exact test with two-sided *P*- value was used to determine statistical significance between two proportions. A *P*-value of less than 0.05 was considered statistically significant.

4.9.   References

1. Groth AC, Olivares EC, Thyagarajan B, Calos MP*, et al*. A phage integrase directs efficient site-specific integration in human cells. Proc Natl Acad Sci U S A. (2000) 97:5995-6000.

2. Preuss E, Treschow A, Newrzela S, Brücher D*, et al*. TK.007: A novel, codon-optimized HSVtk(A168H) mutant for suicide gene therapy. Hum Gene Ther. (2010) 21:929-941.

3. Hockemeyer D, Soldner F, Beard C, Gao Q*, et al*. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat Biotechnol. (2009) 27:851-857.

4. Bi L, Lawler AM, Antonarakis SE, High KA*, et al*. Targeted disruption of the mouse factor VIII gene produces a model of haemophilia A. Nat Genet. (1995) 10:119-121.

5. Lenting PJ, van Mourik JA, Mertens K. The life cycle of coagulation factor VIII in view of its structure and function. Blood. (1998) 92:3983-3996.

6. Miao HZ, Sirachainan N, Palmer L, Kucab P*, et al*. Bioengineering of coagulation factor VIII for improved secretion. Blood. (2004) 103:3412-3419.

7. Swaroop M, Moussalli M, Pipe SW, Kaufman RJ*, et al*. Mutagenesis of a potential immunoglobulin-binding protein-binding site enhances secretion of coagulation factor VIII. J Biol Chem. (1997) 272:24121-24124.

8. Miller JC, Holmes MC, Wang J, Guschin DY*, et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol. (2007) 25:778-785.

9. Guo J, Gaj T, Barbas CF. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. J Mol Biol. (2010) 400:96-107.

10. Doyon Y, Vo TD, Mendel MC, Greenberg SG, *et al*. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat Methods. (2011) 8:74-79.

11. Reza HM, Ng BY, Gimeno FL, Phan TT, *et al*. Umbilical cord lining stem cells as a novel and promising source for ocular surface regeneration. Stem Cell Rev. (2011) 7:935-947.

12. Zhou Y, Gan SU, Lin G, Lim YT, *et al*. Characterization of Human Umbilical Cord Lining Derived Epithelial Cells and Transplantation Potential. Cell Transplant. (2011)

13. Huang L, Wong YP, Gu H, Cai YJ, *et al*. Stem cell-like properties of human umbilical cord lining epithelial cells and the potential for epidermal reconstitution. Cytotherapy. (2011) 13:145-155.

14. Sivalingam J, Krishnan S, Ng WH, Lee SS, *et al*. Biosafety assessment of site-directed transgene integration in human umbilical cord-lining cells. Mol Ther. (2010) 18:1346-1356.

15. Chen NKF, Wong JS, Kee IHC, Lai SH, *et al*. Nonvirally modified autologous primary hepatocytes correct diabetes and prevent target organ injury in a large preclinical model. PloS one. (2008) 3:e1734.

16. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, *et al*. Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol. (2006) 357:28-48.

17. Sclimenti CR, Thyagarajan B, Calos MP. Directed evolution of a recombinase for improved genomic integration at a native human sequence. Nucleic Acids Res. (2001) 29:5044-5051.

18. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. (2001) 25:402-408.

19. Padilla-Nash HM, Barenboim-Stapleton L, Difilippantonio MJ, Ried T*, et al*. Spectral karyotyping analysis of human and mouse chromosomes. Nat Protoc. (2006) 1:3129-3142.

20. Margaritis P, Arruda VR, Aljamali M, Camire RM*, et al*. Novel therapeutic approach for hemophilia using gene delivery of an engineered secreted activated Factor VII. J Clin Invest. (2004) 113:1025-1031.

21. Qiu P, Shandilya H, D'Alessio JM, O'Connor K*, et al*. Mutation detection using Surveyor nuclease. Biotechniques. (2004) 36:702-707.

22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. (2009) 25:1754-1760.

APPENDICES

Appendix 1   Complete list of primers used in this project

Purpose	Forward sequence (5' → 3')	Reverse sequence (5' → 3')	Amplicon size (bp)	Annealing temperature (°C)
Primers for phiC31 integrase study				
Primers for cloning of porcine/human hybrid FVIII cDNA				
Cloning human FVIII A1 – B domains	tgtagcgctagcatg caaatag	gaataaggcgatatctt tagtcaa	3000	60
Cloning human FVIII B – C2 domains	gcaaagcccgggaggactgaa	cagtggctcgaggtcagtagaggt	2100	60
Introduction of F309S mutation to human FVIII	agtttctactgtcttgtcatatctct	agagatatgacaagacagtagaaact	Not applicable	55
Sequencing primer to confirm F309S mutation	gtcttcatgctgttggtg		Not applicable	
Human signal peptide to FVIII porcine A1 domain	atgcaaatagagctctccacctgtttctttctgtgtcttttgcgattctgctttagtgccatcaggagatactacctgggcgcagtggaactgt	aggatgcttcttggcaactgagcggatttggataaagggaga	1104	60
Human FVIII A2 domain to partial B-domain	tctccctttatccaaatccgctcagttgccaagaagcatcct	gcgggggctctgattttcatcctc	1854	60
Porcine A3 domain	agctttcagaagagaacccgacac	tcccaggggagtctgacacttcttgctgtacaccaggaaagt	1164	60
Human C1 to C2 domains	actttcctggtgtacagcaagaagtgtcagactcccctggga	agtgctagctcagtagaggtcctgtgcc	936	60
Primers for RT-PCR analysis of pluripotency genes				
Nanog	ttccttcctccatggatctg	tctgctggaggctgaggtat	159	60
Oct-4	ggttctatttgggaaggtattcag	ggtttctgctttgcatatctc	212	60
Gamma-actin	accactggcattgtcatggactct	atcttgatcttcatggtgctgggc	545	60
Primers for detecting phiC31-mediated transgene integration				
CHOSeq R Sequencing *attB* integrants		tcccgtgctcaccgtgaccac	Not applicable	

8p22 left integration junction PCR	gggctctggagtaa aggtgaaa	gttcgccgggatcaact acc	454	60
8p22 right integration junction PCR	tcgacgatgtaggtc acgg	gcatggcctcatttccg tct	333	60
Primers for RT-PCR of *DLC1* transcript				
DLC1 exon 1-2	tcctgccccaatgga atgtc	gttggtgtgcctgatgg aga	305	62
DLC1 exon 8-9	gaaggggatgcag cggatag	agcagggccgttagct ttag	390	62
GAPDH exon 6 - 7	gcctcctgcaccac caact	cgcctgcttcaccacct tc	348	62
Primers for FISH experiments				
FISH probe for detecting pattB EGFP-C1	ccggtcgccaccat ggtgag	ctgagtccggacttgta cag	740	62
FISH probe for detecting pattB Hhbrid FVIII	ttgcacgcaggttct ccggc	ggcgtcgcttggtcggt cat	817	62
Primers for AAVS1 ZFN study				
Primers for HSV-TK study				
Amplification of MC1 promoter	gtc gcg agt cga gca gtg tgg tt	ggc tag cac gcg ctt cta caa g	379	62
HSV-TK RT-PCR/genomic PCR	gttcgaccaggctg cgcgtt	gtgttgtgtggtgtagat gt	280	62
HSV-TK.007 RT-PCR/genomic PCR	gcatgacccccag gccgtgc	cacgttgtacaggtcgc cg	412	62
Primers for mutagenesis of *Fok*I endonuclease				
E490K mutation	gca acg ata tgt caa aga aaa tca aac acg	cgt gtt tga ttt tct ttg aca tat cgt tgc	Not applicable	55
I538K mutation	cac gat taa atc ata aga cta att gta atg gag c	gct cca tta caa tta gtc tta tga ttt aat cgt g	Not applicable	55
Q468E mutation	cca agc aga tga aat gga acg ata tgt cga ag	ctt cga cat atc gtt cca ttt cat ctg ctt gg	Not applicable	55
I499L mutation	cac gaa aca aac atc tca acc cta atg aat gg	cca ttc att agg gtt gag atg ttt gtt tcg tg	Not applicable	55
S418P mutation	att gaa att gcc aga aat ccc act cag gat aga att ctt	aag aat tct atc ctg agt ggg att tct ggc aat ttc aat	Not applicable	55

K441E mutation	gtt tat gga tat aga ggt gaa cat ttg ggt gga tca agg	cct tga tcc acc caa atg ttc acc tct ata tcc ata aac	Not applicable	55
H537R mutation	cag ctt aca cga tta aat c**gt** aag act aat tgt aat gga	tcc att aca att agt ctt acg att taa tcg tgt aag ctg	Not applicable	55
N496D mutation	gaa gaa aat caa aca cga **g**ac aaa cat ctc aac cct aat	att agg gtt gag atg ttt gt**c** tcg tgt ttg att ttc ttc	Not applicable	55
Primers for RT-PCR of *PPP1R12C* transcript				
PPP1R12C exons 4 - 6	agaggaattgctcct tcatgacac	caacaggctcagtactt cctcatc	322	60
PPP1R12C exons 8 -12	gagctctgtgtgtcg tctg	ccgtggaggctgtggg ga	552	60
GAPDH exon 6 - 7	gcctcctgcaccac caact	cgcctgcttcaccacct tc	348	60
Primers for RFLP assay, *Cel*-1 assay and integration junction PCR				
AAVS1 ZFN left RT-PCR	gcagacaggccctg gaca	cccaggttgtttctctg	180	62
AAVS1 ZFN right RT-PCR	agaaacttcatcctg caga	gtcagctggtcgtgtct	180	62
Integration junction PCR to detect AAVS1 ZFN integration of 50-bp donor DNA	agcttgaattctctag aaatattctcgaggtt taaacgtcgacgc	ggaacggggctcagtc tg	1026	62
RFLP PCR to detect AAVS1 ZFN integration of 50-bp donor DNA	ggccctggccattgt cactt	ggaacggggctcagtc tg	1880	62
Surveyor mutation detection PCR (*Cel*-1 assay)	ttcgggtcacctctc actcc	ggctccatcgtaagca aacc	514	60
Control PCR to amplify AAVS1 genomic region	aagaagcgcaccac ctccaggttctc	atgacctcatgctcttg gccctcgta	976	60
Left integration junction PCR to detect integration of 4-kb GFP donor DNA	ggccctggccattgt cactt	cgtcaataggggggcgt acttggcatatgatac	1237	62
Right integration	gacatagcgttggct acccgtgatattgct	ggaacggggctcagtc tg	1356	62

junction PCR to detect integration of 4-kb GFP donor DNA	gaagagc			
Left integration junction PCR to detect integration of 9-kb hybrid FVIII donor DNA	ggccctggccattgtcactt	cgtcaatagggggcgtacttggcatatgatac	1237	62
Right integration junction PCR to detect integration of 9-kb hybrid FVIII donor DNA	gacatagcgttggctacccgtgatattgctgaagagc	ggaacggggctcagtctg	1356	62
Left integration junction PCR to detect AAVS1-SA-2A-Puro-pA integration	ggccctggccattgtcactt	cggtcatctcgagcctaggg	1103	62
Right integration junction PCR to detect AAVS1-SA-2A-Puro-pA integration	tcaccgagctgcaagaact	ggaacggggctcagtctg	1621	62
PCR to detect complete integration of AAVS1-SA-2A-Puro-pA	ctggccattgtcacttgcg	ggaacggggctcagtctg	3084	62
Left integration junction PCR to detect AAVS-CAGGS-EGFP integration	ctggccattgtcacttgcg	cggtcatctcgagcctaggg	1099	62
Right integration junction PCR to detect AAVS-CAGGS-EGFP integration	ctactcccagtcatagctgtc	cttggccacgtaacctgaga	1370	62
Genomic PCR to detect left half of AAVS-CAGGS-EGFP integration	ctggccattgtcacttgcg	gatggggagagtgaagcagaacg	2382	60
Genomic PCR to detect right	aaacggccacaagttcagcg	cttggccacgtaacctgaga	2517	60

half of AAVS-CAGGS-EGFP integration				
Left integration junction PCR to detect SA-2A Puro hybrid FVIII integration	ctggccattgtcacttt gcg	cggtcatctcgagccta ggg	1099	62
Right integration junction PCR to detect SA-2A Puro hybrid FVIII integration	actttcctggtgtaca gcaagaagtgtcag actcccctggga	cttggccacgtaacctg aga	3187	60
Genomic PCR to detect left half of SA-2A Puro hybrid FVIII integration	ctggccattgtcacttt gcg	gcgggggctctgatttt catcctc	6859	60
Genomic PCR to detect right half of SA-2A Puro hybrid FVIII integration	agctttcagaagaga acccgacac	cttggccacgtaacctg aga	4156	60
Primers to verify transcriptome data by RT-PCR				
LILRB4	ggacattggcccag agacag	gtcttcatcgtgtgggct ct	197	60
ISOC2	tgacggagcagtac ccacaag	ggataagaacggggg tccaagatg	196	60
PPP6R1	gcgctacaagtacc ccagtg	tccgaagaaaggacac gagc	215	60
NAT14	gctcccgaaaccttg tcgaa	ccttcacgccggccttc	175	60
ZNF579	aaggaggcgaggc atggat	gtaggggaaacggaa gaggc	184	60
FIZ1	cagagggaggtag agagccc	ctgtgcttgaaaccctt gcc	235	60
RDH13	ggtccccttcccagc tgaa	atgatgttgcctcctctc ctg	278	60
DUSP1	ggatacgaagcgttt tcggc	ggccaccctgatcgta gagt	151	60
DUSP6	acctggaaggtggc ttcagta	accatccgagtctgttg cac	197	60
CDC6	agaagggcccccatg attgtg	tgcagcattgtccagaa cct	296	60
DUSP16	ctgcttgcaggtggg tttg	gcacacgcaggaaat gagac	271	60
Primers for analysis of potential off-target sites of AAVS1 ZFN				

AAVS1 locus	ttcgggtcacctctc actcc	ggctccatcgtaagca aacc	469	60
OT1	tttaagaactgtaacc tattttccaaagtgttt g	cag cct ggc caa cat ggt gaa ac	389	62
OT2	aaggtgtaagtgga gccacaaggct	tgtggtccttgctggatc aggaa	308	60
OT3	ttggaaataagaccc atttgttgatgaga	ctggctcattccaacgt ccatgt	389	58
OT4	gacttggtggttggc agaatacacc	gggtaaggtcagatag ggctgtaagactc	601	60
OT5	ggaacaaggcacct ggctcc	ccattcccgggagaaa tctc	353	58
OT6	tgagtttgggcctga ggtcatc	ggcttggaaacaccca ggtg	320	60
OT7	ctttgagtttagcagc ttccaggaacc	gttttatcttcataaggta gtgggcagatgg	631	60
OT8	ggtcctcaccccatc ttcatc	aaagagagggctggt gaggc	375	60
OT9	gttgcgagagtccct actgg	agcctgaagttgagcct gtc	348	60
OT10	cacagagttcaggg gatcgt	gccactttgtattgggtg gt	650	60

Appendix 2   Plasmids used in this project

Plasmids used in PhiC31 integrase study
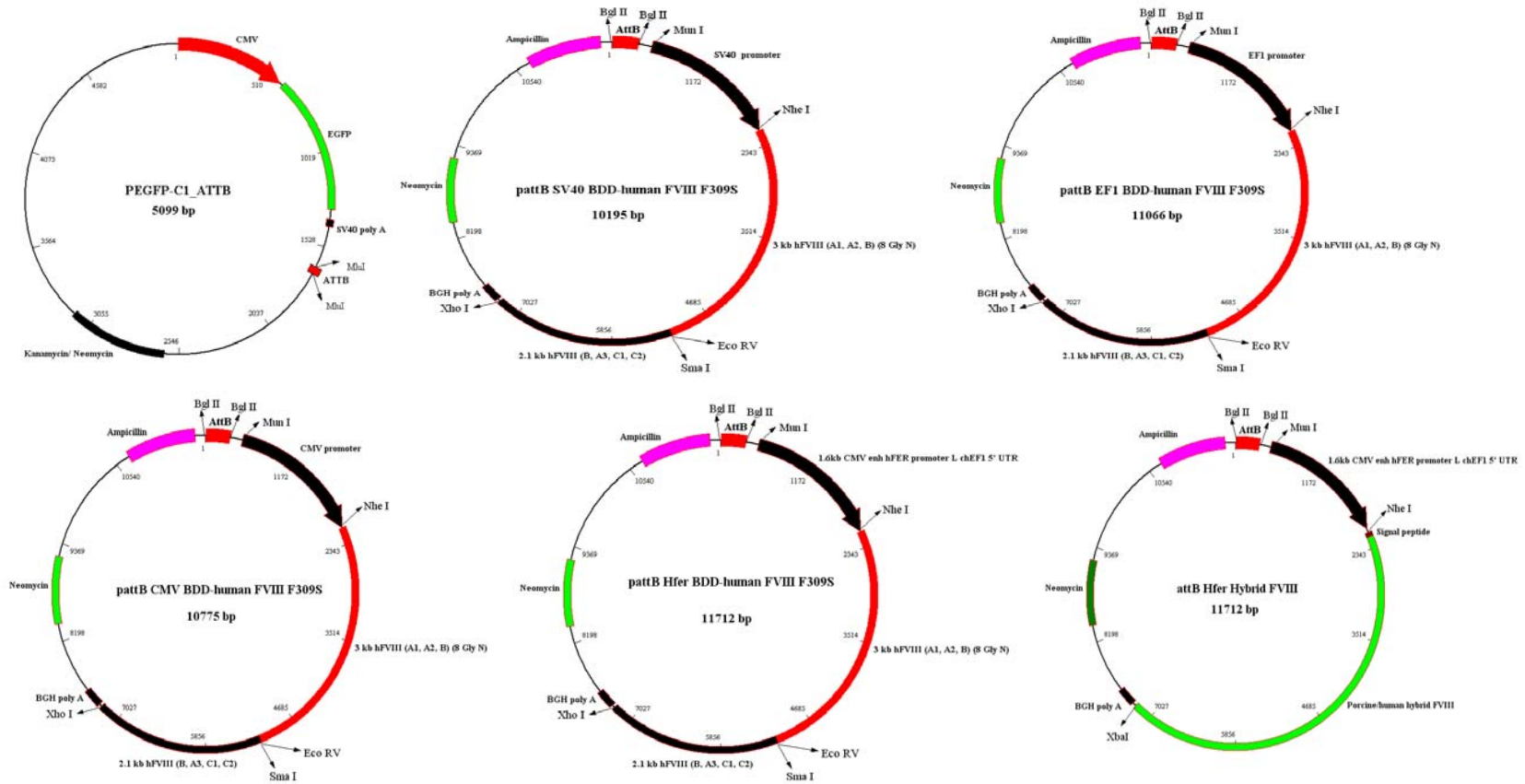
Plasmids used in the AAVS1 ZFN study

AAVS1 ZFN plasmids
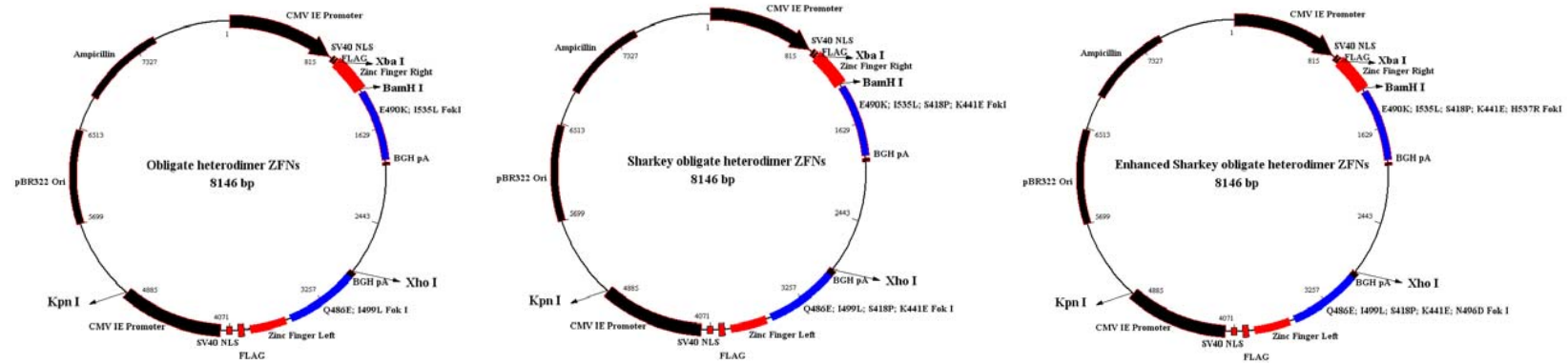
Donor plasmids for AAVS1 ZFN study
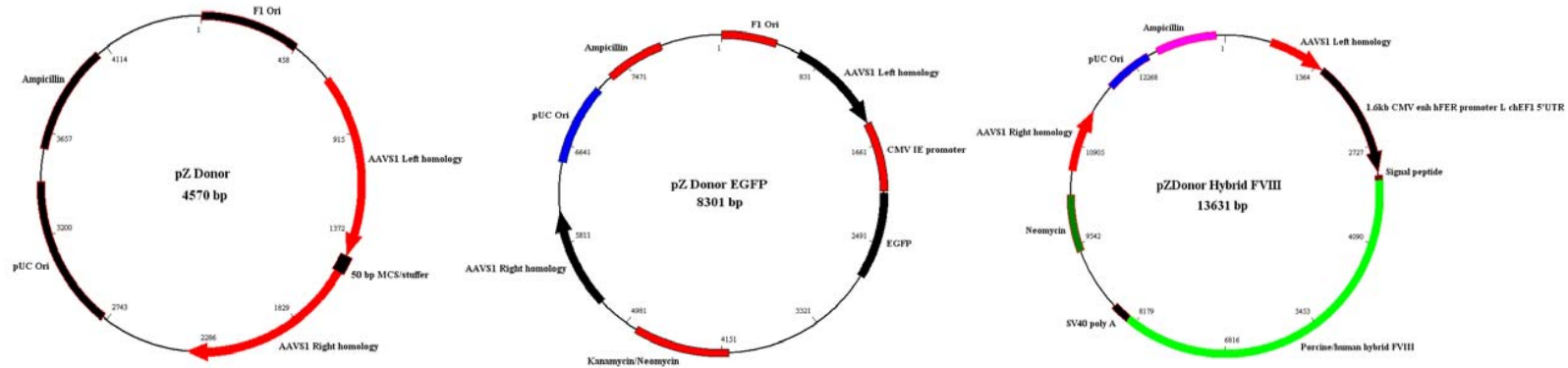
Donor plasmids (promoterless puromycin resistance cDNA) for AAVS1 ZFN study
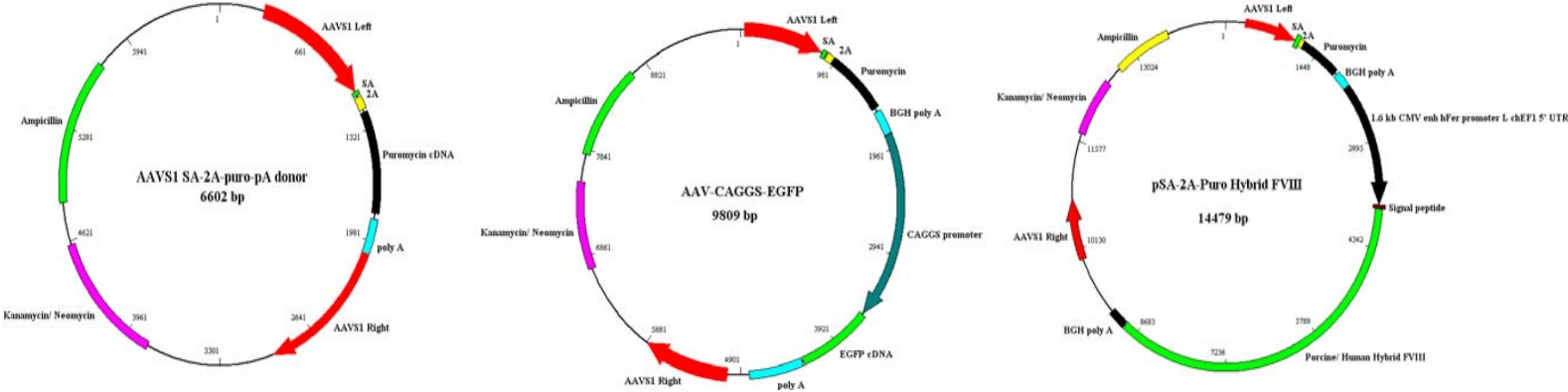
Plasmids used in the HSV-TK007 study
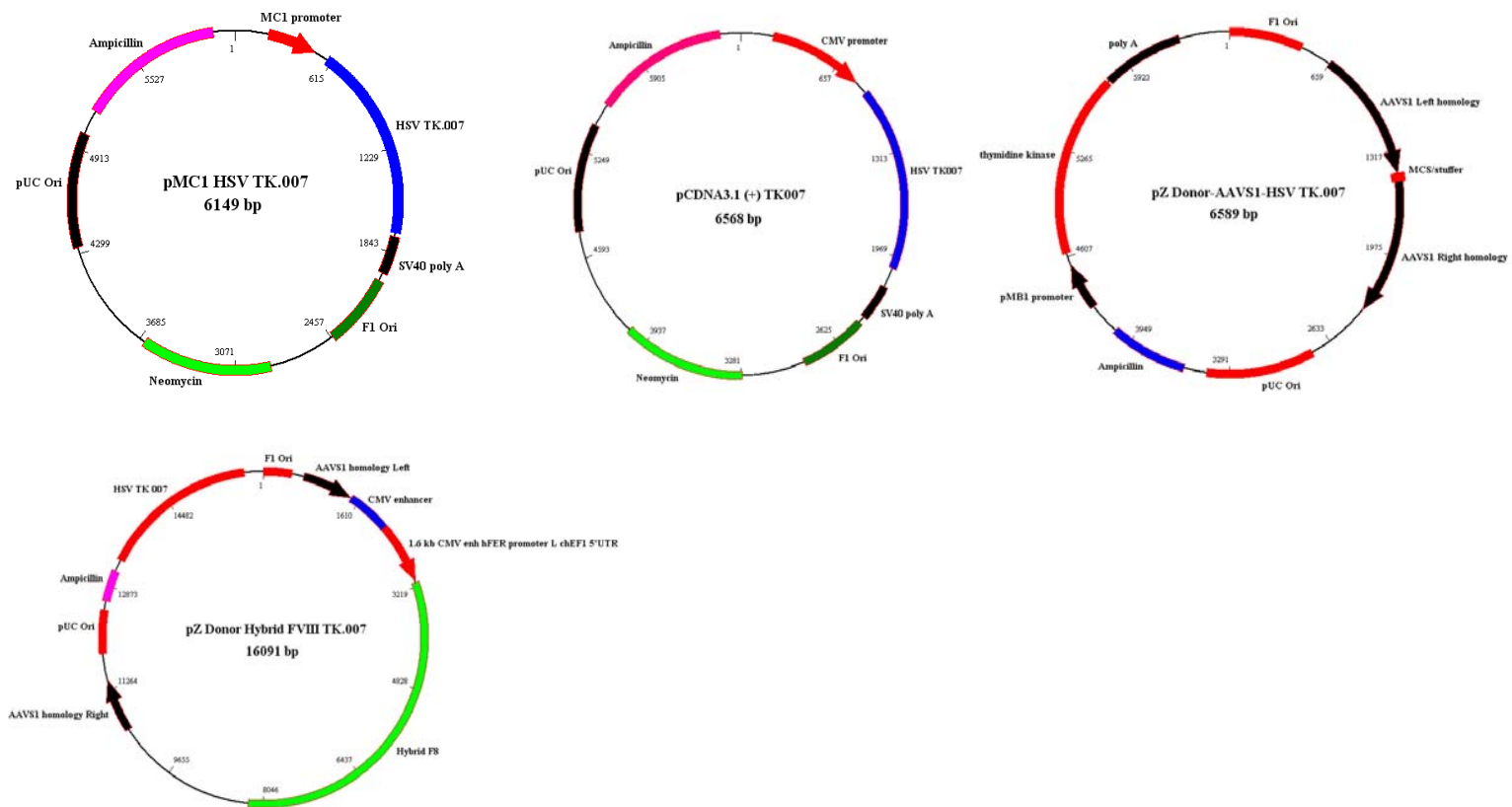
Appendix 3   Cloning strategy for human/porcine hybrid FVIII

Step 1: A1-PCR (Template: porcine mRNA; Primers: A1F, A1R) (1104 bp)

Step 2: A2-B PCR (Template: pSP64-F8 plasmid; Primers: A2F, B domain R) (1854)

Step 3: A1 + A2-B Overlap PCR (Template: A1 and A2/B amplicons; Primers: NheIF, B domain R) (2958 bp) **(Stratablunt cloning)**

Step 4: A3 PCR (Template: porcine mRNA; Primers: A3F, A3R) (1164 bp)

Step 5: C1-C2 PCR (Template: pSP64-F8 plasmid; Primers: C1F, C2R) (936 bp)

Step 6: A3 + C1-C2 Overlap PCR (Template: A3 and C1/C2 amplicons; Primers: A3F, C2R) (2100 bp) **(Stratablunt cloning)**

Step 7: (A1-A2-B) + (A3-C1-C2) Blunt Ligation (5058 bp)

Step 8: Blunt clone into pAttB Hfer BDD-human FVIII F309S

Appendix 4A   Workflow for identifying phiC31 integrase-mediated integration events
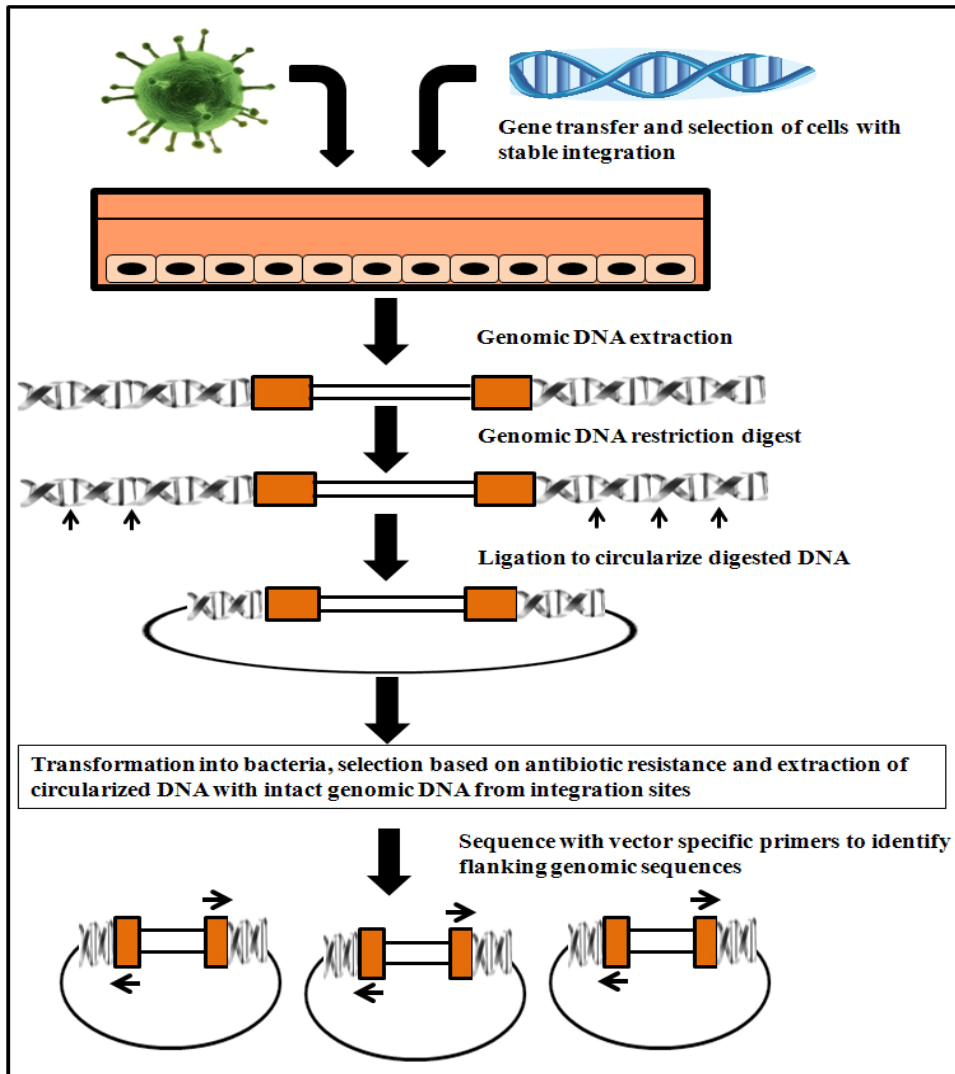
Appendix 4B   Workflow for identifying phiC31 integrase-mediated integration events
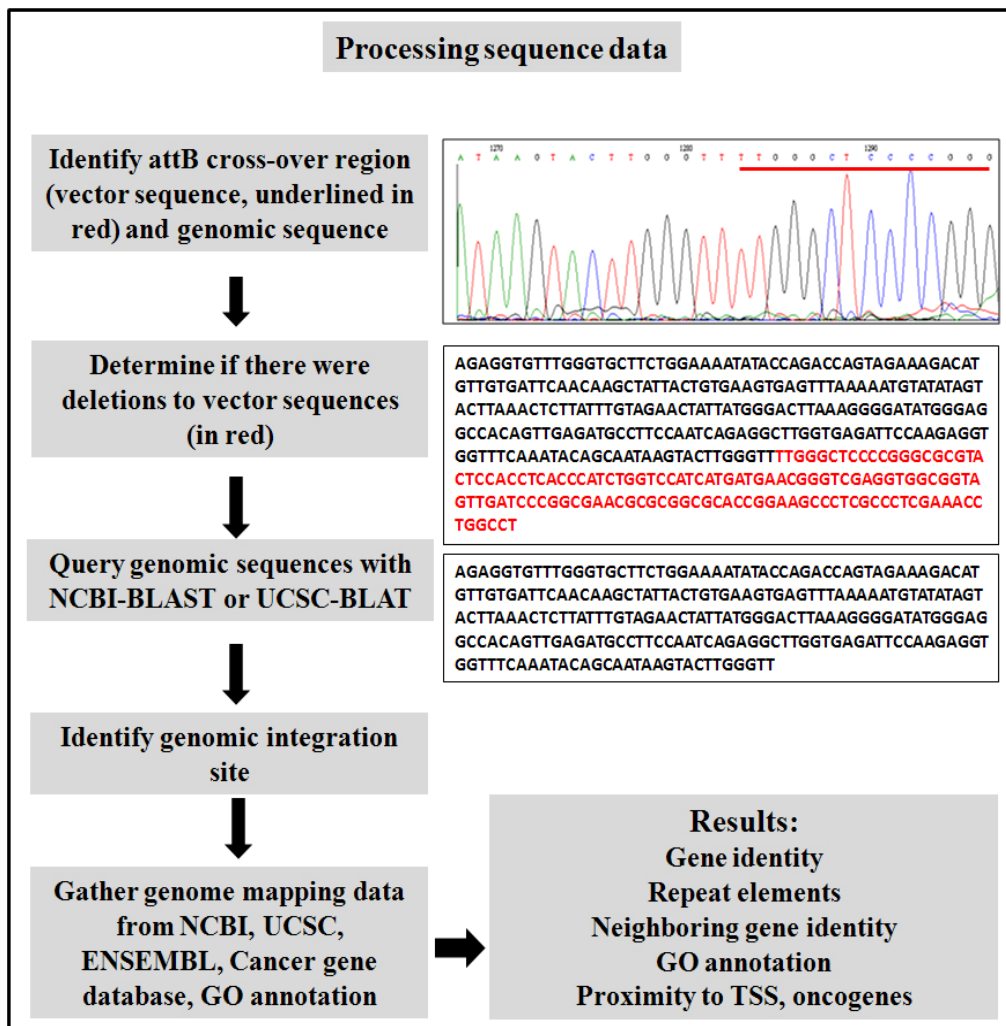
Appendix 5   Sequences of integration junction amplicons for phiC31 integrase- modified 8p22 oligoclones

AttB sequence

Cross-over region

TGGAGATCtgtcgacgatgtaggtcacggtctcgaagccgcggtgcgggtgcc**agggcgtgccct**tgggctccccgggcgcgtactccac
ctcacccatctggtccatcatgatgaacgggtcgaggtggcggtagttgatcccggcgaacgcgcggcgcaccgggaagccctcgccctcgaa
accgctgggcgcggtggtcacggtgagcacgggacgtgcgacggcgtcggcgggtgcggatacgcggggcagcgtcagcgggttctcgacg
gtcacggcgggcatgtcgacagccaagccagatcTCCCGATCCGTC

Sequence of 8p22 integration junctions

Left integration junction

GGGCTCTGGAGTAAAGGTGAAAGGCAGGTTACCCTTCAACCTAAAGCCCTTGTTTGCA
TGTTTTTTTGTTGTTCTGTACTGTAGTTGAGAGAGCAAAGACTAAGCAAAGTTTAGAG
CATTTCTACATTAAAAGAGGTGTTTGGGTGCTTCTGGAAAATATACCAGACCAGTAGA
AAGACATGTTGTGATTCAACAAGCTATTACTGTGAAGTGAGTTTAAAAATGTATATAGT
ACTTAAACTCTTATTTGTAGAACTATTATGGGACTTAAAGGGGATATGGGAGGCCACAG
TTGAGATGCCTTCCAATCAGAGGCTTGGTGAGATTCCAAGAGGTGGTTTCAAATACAG
CAATAAGTACTTGGGTTTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTG
GTCCATCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAAC...............
....................vector
sequence......................
TCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCC
CTTCCCTTGGTGTCCCCATGGAGATTTTAAGCCATGACGCAATGTTTAAATCAGAGTG
GTATTTTTATGACTTAAGCGGGTAAATATGCAATTGGAAAATATTCAGGGAAGGGTGAT
TTGGTCCAGAAGAGTGGGGGCATCCAGAGTACAGTGGGTGAAATGGATCGGACTTTT
TGGAAGAGAGCCTTGTGCTGGACAGGATGGTCCAGTATTGTCAACACAAGTTTCTCAT
GCTTCACTCTCCTTCCTAGCAACAGGAAGACGGAAATGAGGCCATGC

Right integration junction

Deletion of 7 bp of attB vector at left junction

Genomic DNA
attB vector sequence
Genomic DNA

Deletion of 7 bp of genomic DNA at right junction

Left integration junction

Right integration junction
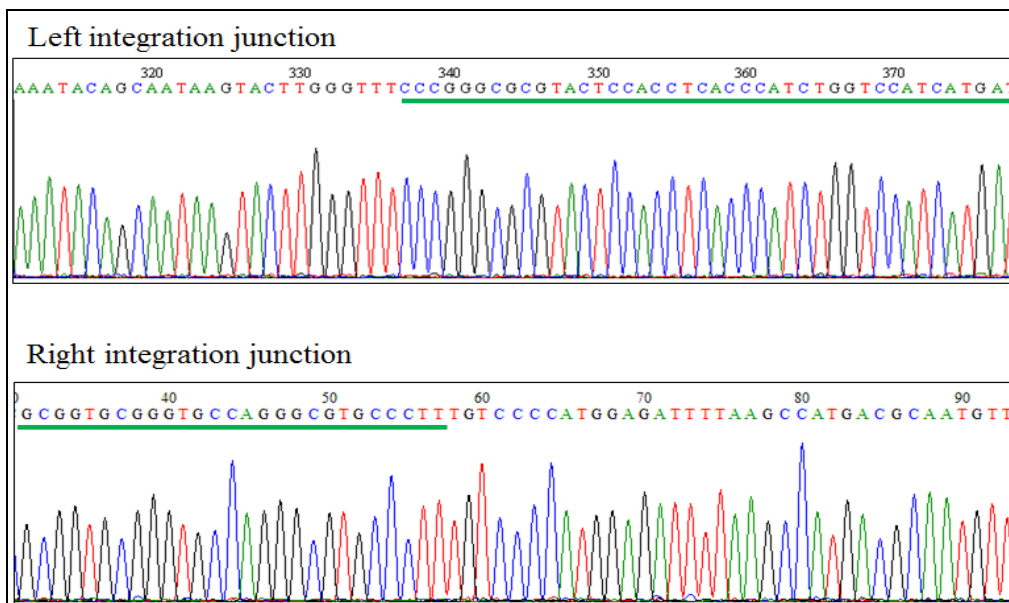
Top: AttB vector sequence highlighted to show cross-over region. **Middle**: Illustration of genomic sequence (in green) with integrated attB vector sequence (in red). **Bottom**: DNA sequence chromatograms of left and right integration junction PCR products showing genomic sequence and vector sequences (underlined in green).

Appendix 6   Sequences of AAVS1 integration junction amplicons for pSA-2A-puro hybrid FVIII stables
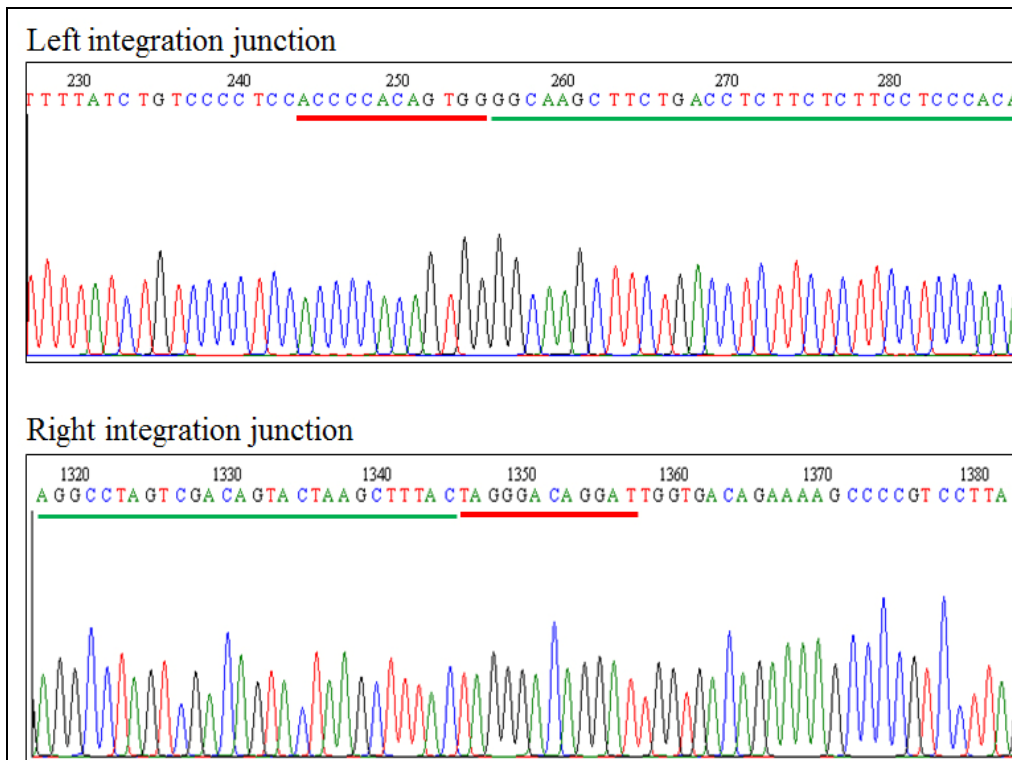
Left integration junction

```
      230        240        250        260        270        280
TTTTATC TG TCCCC TCCACCCCACAG TG GGGC AAGC TTC TG ACC TC TTC TC TTCC TCCCACA
```

Right integration junction

```
    1320       1330       1340       1350       1360       1370       1380
A GGC CTA GT CGA CA GTA CTAA G CTT TA C TA GGGA CA GGA TTG GTGA CA GAAAA G CCCCGT CCTTA C
```

DNA sequence chromatograms of left and right integration junction PCR products showing genomic DNA sequences, AAVS1 ZFN recognition half sites (underlined in red) and vector sequences (underlined in green).

356

Appendix 7   **List of genes that were significantly altered (≥2.5-fold) in all 8 oligoclonal CLECs with transgene integration at 8p22 locus compared to unmodified CLECs**

GENE ONTOLOGY CLASSIFICATIONS	OVEREXPRESSED GENES
Apoptosis	*BCL2A1*, CLU, FOXQ1, TNFRSF21
Cell cycle	*PLCB1*
Cell adhesion	*COL7A1*, NEGR1
Cellular transport	KCNE4, NALCN, SCN2A, SLC12A8, SLC22A3, SLC46A3, SLC7A8
Differentiation and development	CHN2, DCN, HOXD10, MAB21L1, PDZRN3, RBP4, SHOX2, SPAG4, SULF1
Immune response	ADSSL1, ALOX5AP, C3, HLA-B, IFI44L, LY96, SAA1, SAA2-SAA4, SAMSN1, TNFAIP3
Metabolism	ABCA8, ABCC3, *ADAM12*, AKR1C1, AKR1C2, C1R, C1S, CHI3L2, COX7A1, *GALNT5*, HSD11B1, HSD17B2, KYNU, PAMR1, PAPPA, PGM2L1, PLCB4, TTC3, TYRP1, XYLT1
Signal transduction	ANGPTL1, CXCL6, DEPTOR, *FGF7*, *FGF11*, GPR155, INSR, IQGAP2, ITGA1, LRRK2, *MAP3K8*, MYOCD, *PDGFRL*, PDZRN4, PELI2, QPCT, SORBS2, SPOCK3, ST6GALNAC5
Others	ACTA2, BCHE, C10orf10, C4orf47, CCDC102B, CHN2, F8, FAM117B, HTRA3, KIAA1324L, LOC100129518, LOXL4, NCKAP5, P4HA3

GENE ONTOLOGY CLASSIFICATIONS	UNDEREXPRESSED GENES
Apoptosis	C11orf82, KCNMA1, LMNB1, PEG10, PMAIP1, TPX2
Cell cycle	ANLN, ASPM, AURKA, AURKB, BIRC5, *BLM*, *BUB1*, *BUB1B*, C15orf23, C2orf18, *CASC5*, CCNA1, *CCNA2*, *CCNB1*, CCNB2, *CCNF*, CDC20, *CDC25A*, *CDC25C*, CDC45, CDC6, CDC7, CDCA3, CDCA5, CDCA8, CDK1, *CDKN3*, CENPE, CENPF, CENPI, CENPK, CENPM, CENPN, CENPP, CENPV, CEP, CKAP2, CKS1B, DHFR, DLGAP5, DNA2, ERCC6L, ESPL1, EZH2, FAM83D, *FBX05*, FEN1, FOXM1, GINS1, GINS2, GINS4, GPSM2, GTSE1, HAUS8, HJURP, IQGAP3, KIF18A, KIF20A, KIF23, KIF2C, KIFC1, KNTC1, LIN9, *MAD2L1*, MASTL, MIS18A, MELK, MLF1IP, NCAPG, NCAPG2, NCAPH, NDC80, NEK2, NUSAP1, OIP5, PLK4, POLA1, POLA2, POLE2, PRC1, PRIM1, RACGAP1, RRM2, SASS6, SGOL1, SGOL2, SKA1, SKA3, SMC4, SPAG5, SPC24, *TOP2A*, TYMS, UBE2C, ZWILCH, ZWINT
Cell adhesion	CLDN11, ITGBL1, TMSB15A, TROAP
Cellular transport	KIF4A, KIF4B, KIF11, KIF15, SLC39A4, SLC8A1, TMEM48, TRPV2, *VAMP8*
DNA repair/replication	*BRAC1*, *BRAC2*, BARD1, CDCA7, CDT1, CHAF1A, CHAF1B, CLSPN, DBF4, DBF4B, DSCC1, DTL, EXO1, *FANCA*, *FANCD2*, FANCI, GAS2L3, GINS3, MCM10, MCM2, MCM3, MCM4, MCM7, MCM8, MMS22L, MND1, MNS1, NEIL3, PARPBP, POLQ, PSMC3IP, RAD18, RAD51, RAD51AP1, RAD54L, RFC2, RFC3, RFC4, RFC5, RNASEH2A, TIPIN, TRIP13, *UCHL5*, UHRF1
Differentiation and development	BEX1, CCBE1, FIGNL1, HELLS, *STIL*, STMN1, TACC3, *TIMP3*

GENE ONTOLOGY CLASSIFICATIONS	UNDEREXPRESSED GENES
Immune response	CD302, TNFSF4
Metabolism	APOBEC3A, APOBEC3B, ATAD2, *ATP8B1*, BACE2, BORA, CPS1, FAR2, GPAT2, HAS1, MKI67, NAGS, *TK1*
Signal transduction	ARHAP11A, CARD10, DEPDC1B, FAM129A, *FLT1*, HMGB2, MOK, PBK, PIF1, PMEPA1, PPM1F, *PTPRD*, RASIP1, RGS4, TTK, *ST6GAL1*
Transcription	ASF1B, DDX39A, DEPDC1, DNAJC9, E2F7, HOXB6, IRX5, KHDRBS3, *MYBL1*, PHF19, PRRX1, *SMYD3*, TBX18, TCF19, TIMELESS, TMPO, *WHSC1*
Others	C18orf54, C1orf115, C1orf135, C4orf46, CEP128, CKAP2L, DLEU2, DLEU2L, FAM54A, FAM64A, FAM72A, FRMF3, HMMR, KIAAO101, KIAAO0408, KIAA1524, KLHL23, KRTAP2-4, LETM2, LOC100509445, LY6K, PAQR4, PRR11, PTGFRN, S100A16, SAPCD2, SHCBP1, SLITRK4, TMEM106C, TRIM59, WDHD1, WDR76

Gene Ontology classification of genes, common to all 8 oligoclonal CLECs with transgene integration at 8p22 locus, whose expression was ≥ 2.5-fold different compared to wild type CLECs. Transcriptome profiling (Affymetrix Primeview array) was performed with oligoclonal CLECs 2 months after phiC31 integrase-mediated stable transgene integration. Potential oncogenes and tumor suppressor genes are underlined and italicized.

Appendix 8   KEGG pathways of significantly down-regulated genes common to all 8 oligoclonal CLECs with transgene integration at 8p22 locus

Functional Annotation Chart

Help and Manual

Current Gene List: List_2
Current Background: Homo sapiens
245 DAVID IDs

Options

Rerun Using Options   Create Sublist

10 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
☐	KEGG_PATHWAY	DNA replication	RT		15	6.1	1.7E-18	7.5E-17
☐	KEGG_PATHWAY	Cell cycle	RT		21	8.6	1.1E-17	2.4E-16
☐	KEGG_PATHWAY	Oocyte meiosis	RT		11	4.5	8.5E-7	1.3E-5
☐	KEGG_PATHWAY	Progesterone-mediated oocyte maturation	RT		9	3.7	9.9E-6	1.1E-4
☐	KEGG_PATHWAY	Mismatch repair	RT		5	2.0	1.7E-4	1.5E-3
☐	KEGG_PATHWAY	Pyrimidine metabolism	RT		7	2.9	1.1E-3	8.5E-3
☐	KEGG_PATHWAY	p53 signaling pathway	RT		6	2.4	1.5E-3	9.8E-3
☐	KEGG_PATHWAY	Nucleotide excision repair	RT		5	2.0	2.1E-3	1.2E-2
☐	KEGG_PATHWAY	Homologous recombination	RT		4	1.6	5.0E-3	2.5E-2
☐	KEGG_PATHWAY	Base excision repair	RT		3	1.2	7.1E-2	2.8E-1

199 gene(s) from your list are not in the output.

List of KEGG pathways derived from DAVID analysis of significantly down-regulated genes common to all 8 oligoclonal CLECs with transgene integration at 8p22 locus

Appendix 9   Evaluation of selected genes derived from transcriptome study of transgenic CLECs with targeted integration of Puro hybrid FVIII cassette at AAVS1 locus following ZFN treatment

Gene symbol	Change in expression compared to control CLECs
Neighbouring genes within 1-Mb of AAVS1 locus	
BRSK1	Not expressed
C19ORF51 (DNAAF3)	Not expressed
COX6B2	Not expressed
EPS8L1	Not expressed
FAM71E2	Not expressed
FCAR	Not expressed
FIZ1	No change
GP6	Not expressed
HSPBP1	Not expressed
ISOC2	No change
KIR2DL1	Not expressed
KIR2DL3	Not expressed
KIR2DL4	Not expressed
KIR2DS4	Not expressed
KIR3DL1	Not expressed
KIR3DL2	Not expressed
KIR3DL3	Not expressed
LILRA1	Not expressed
LILRA2	Not expressed
LILRB1	Not expressed
LILRB4	Up-regulated 1.6-fold
LILRP2	Not expressed
NAT14	Down-regulated 1.4-fold
NCR1	Not expressed
NLRP2	Not expressed
NLRP7	Not expressed

PPP6R1	Down-regulated 1.3-fold
PTPRH	Not expressed
RDH13	Down-regulated 1.6-fold
SBK2	Not expressed
SGK110	Not expressed
SHISA7	Not expressed
SSC5D	Not expressed
SUV420H2	Not expressed
SYT5	Not expressed
TMEM150B	Not expressed
TMEM190	Not expressed
TMEM238	Not expressed
TMEM86B	Not expressed
TNN13	Not expressed
TNNT1	Not expressed
ZNF524	Not expressed
ZNF579	No change
Predicted interacting partners of *PPP1R12C* as determined by Gene Network Central™ and String 9.05	
BRAF	No change
CDC42BPA	No change
CDC45	Not expressed
CDC5L	No change
CDC6	Up-regulated 2-fold
CDC7	Not expressed
DUSP1	Down-regulated 2.4-fold
DUSP16	Down-regulated 5.3-fold
DUSP6	Up-regulated 5.5-fold
GMNN	No change
IL16	No change
MAPK3	No change

MCM10	Not expressed
MCM2	No change
MPRIP	No change
MYL2	Not expressed
PPP1CB	No change
PRKCE	No change
PRKCI	No change
RAF1	No change
SMAD3	No change
STK35	No change
TNKS2	No change
WIPF1	No change
Predicted downstream effector genes identified from the literature	
CDC42	Down-regulated 2.5-fold
CDC42BPA	No change
CDC42BPB	No change
CDC42BPG	Not expressed
CDC42EP1	Not expressed
CDC42EP2	No change
CDC42EP3	No change
CDC42EP4	Down-regulated 2.3-fold
CDC42EP5	No change
CDC42SE1	No change
CDC42SE2	No change
EZR	No change
MYH1	Not expressed
MYH10	No change
MYH11	Not expressed
MYH13	Not expressed
MYH14	Not expressed
MYH15	Up-regulated 14-fold

MYH16	No change
MYH2 /// MYH4	Not expressed
MYH3	Not expressed
MYH6	Not expressed
MYH7	Not expressed
MYH7B	Not expressed
MYH8	Down-regulated 7.7-fold
MYH9	No change
MYL1	Not expressed
MYL10	No change
MYL12A	Up-regulated 3-fold
MYL12B	No change
MYL2	Not expressed
MYL3	Not expressed
MYL4	No change
MYL5	No change
MYL6	No change
MYL6B	No change
MYL7	Not expressed
MYL9	No change
MYLIP	Down-regulated 3.6-fold
MYLK	No change
MYLK2	Not expressed
MYLK3	Not expressed
MYLK4	Not expressed
RDX	No change
RHO	Not expressed
RHOA	No change
RHOB	No change
RHOBTB1	No change
RHOBTB2	No change
RHOBTB3	No change

RHOC	No change
RHOD	Not expressed
RHOF	No change
RHOG	No change
RHOH	Not expressed
RHOJ	No change
RHOQ	No change
RHOT1	No change
RHOT2	No change
RHOU	No change
RHOV	Not expressed
ROCK1	No change
ROCK2	No change
Members of the protein phosphatase family	
PPP1CA	No change
PPP1CB	No change
PPP1CC	No change
PPP1R10	No change
PPP1R11	No change
PPP1R12A	No change
PPP1R12B	No change
PPP1R12C	Down-regulated 2.3-fold
PPP1R13B	Not expressed
PPP1R13L	No change
PPP1R14A	Not expressed
PPP1R14B	No change
PPP1R14C	Not expressed
PPP1R14D	Not expressed
PPP1R15A	No change
PPP1R15B	No change
PPP1R16A	No change
PPP1R16B	Not expressed

PPP1R17	Not expressed
PPP1R18	No change
PPP1R1A	Not expressed
PPP1R1B	No change
PPP1R1C	Not expressed
PPP1R2	No change
PPP1R21	No change
PPP1R26	No change
PPP1R27	No change
PPP1R2P9	Not expressed
PPP1R32	Not expressed
PPP1R35	No change
PPP1R36	Down-regulated 2.7-fold
PPP1R37	Not expressed
PPP1R3A	Not expressed
PPP1R3B	No change
PPP1R3C	No change
PPP1R3D	Not expressed
PPP1R3E	No change
PPP1R3F	Not expressed
PPP1R7	No change
PPP1R8	No change
PPP1R9A	Not expressed
PPP1R9B	Not expressed
PPP2CA	No change
PPP2CB	No change
PPP2R1A	Not expressed
PPP2R1B	No change
PPP2R2A	Up-regulated 2.2-fold
PPP2R2B	No change
PPP2R2C	No change
PPP2R2D	No change

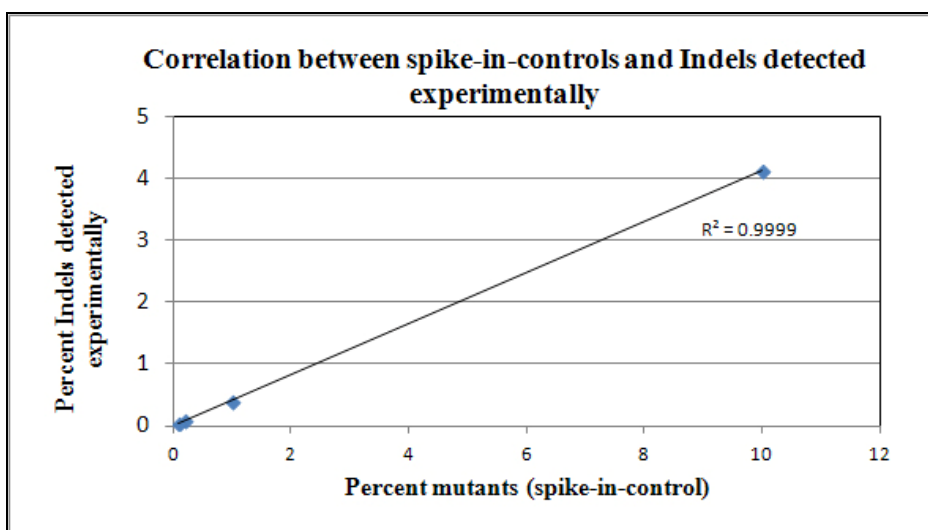|	
PPP2R3A	No change
PPP2R3B	Not expressed
PPP2R3B-AS1	Not expressed
PPP2R3C	No change
PPP2R4	No change
PPP2R5A	Down-regulated 2.7-fold
PPP2R5B	No change
PPP2R5C	No change
PPP2R5D	No change
PPP2R5E	No change
PPP3CA	No change
PPP3CB	No change
PPP3CC	No change
PPP3R1	No change
PPP3R2	Not expressed
PPP4C	No change
PPP4R1	No change
PPP4R1L	Not expressed
PPP4R2	Up-regulated 3-fold
PPP4R4	Up-regulated 2.7-fold
PPP5C	Not expressed
PPP6C	No change
PPP6R1	No change
PPP6R2	No change
PPP6R3	No change

Appendix 10　Top-10 potential off-target sites for AAVS1 ZFNs

Site	Score	Chromosomal location	SELEX predicted sequence	Forward primer	Reverse primer	Amplicon size (bp)
OT1	1.47E-09	chr8:141507033-141507064	GcTCCTGGCCCagTGCTG GCCACTGTGGGTGC	tttaagaactgtaacctatttt ccaaagtgtttg	cag cct ggc caa cat ggt gaa ac	389
OT2	8.39E-10	chr10:47635366-47635397	ACACCCACAGgGGCAGG GGcAGGGCCAGGAcT	aaggtgtaagtggagccac aaggct	tgtggtccttgctggatca ggaa	308
OT3	5.76E-10	chr4:3303431-3303461	TTTCCTGTCCtTtACCTGC CACTGTGGGTtT	ttggaaataagacccatttgt tgatgaga	ctggctcattccaacgtcc atgt	389
OT4	2.67E-10	chr10:117758694-117758724	TCACCCACAGatTTGTAA TAGGGACAGGATT	gacttggtggttggcagaat acacc	gggtaaggtcagataggg ctgtaagactc	601
OT5	1.93E-10	chr9:138563409-138563439	GCACCCACAGcGcAGTGC cAGGGCCAGGAAC	ggaacaaggcacctggct cc	ccattcccgggagaaatct c	353
OT6	1.24E-10	chr14:101033117-101033148	GgTCCTGTCCCTgTGGGA CCCACaGTGGGgGC	tgagtttgggcctgaggtca tc	ggcttggaaacacccagg tg	320
OT7	1.13E-10	chr7:50670960-50670990	GTcCCTGTCCCTATATCC ACACTGTGGcTGG	ctttgagtttagcagcttcca ggaacc	gttttatcttcataaggtagt gggcagatgg	631
OT8	9.84E-11	chr16:28996789-28996820	CATCCTGGCCaTgTTGAT GgCACTGTGtGTGC	ggtcctcaccccatcttcat c	aaagagagggctggtga ggc	375
OT9	8.89E-11	chr19:1224727-1224758	TTTCaTGaCCCTgCTAAGC CCACTGTGGGTGG	gttgcgagagtccctactg g	agcctgaagttgagcctgt c	348
OT10	6.35E-11	chr12:50284986-50285017	CCACCCACAGgGcAGCC AGgAGGGACAGGATG	cacagagttcaggggatcg t	gccactttgtattgggtggt	650

Potential off-target sites, OT1 to OT10 were ranked according to their similarities to the consensus ZFN target site determined by SELEX. Score for each SELEX predicted sequence was derived from experimentally determined base-frequency matrices (Hockemeyer *et al*., 2009).

Appendix 11 Detection sensitivity of deep sequencing: Correlation between spike-in controls (mutant amplicons) and experimentally determined frequency of indels

Ratio of mutant: wt amplicon	Percent mutant (%) (spike-in control)	Total reads mapped	Indels (frequency)	Percent Indels (%)
1:10	10	60562	2504	4.13
1:100	1	55315	222	0.40
1:500	0.2	36029	40	0.11
1:1000	0.1	79659	45	0.056

Correlation between spike-in-controls and Indels detected experimentally

$R^2 = 0.9999$

(**Top**) Table showing ratio and percentage of mutant amplicons to wild-type amplicons, the number of experimentally retrieved mapped reads, and the number and percentage of indels detected in these mapped reads for each spike-in concentration. (**Bottom**) A high linear correlation ($R^2 = 0.999$) was observed between the actual percentage of indels in the spike-in controls and percentage of indels that were experimentally determined by deep sequencing.