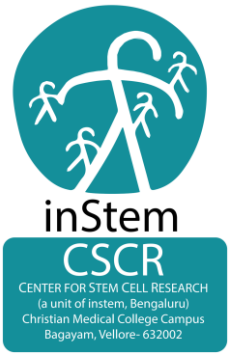
GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS

NITHIN SAM. R

PhD THESIS

2023



SREE CHITRA TIRUNAL INSTITUTE FOR MEDICAL SCIENCE AND TECHNOLOGY, THIRUVANATHAPURAM

An Institution of National Importance established by an Act of the Indian Parliament (Act No.52 of 1980)

Dept. of Science and Technology, Govt. of India

www.sctimst.ac.in

GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS

A THESIS PRESENTED BY

NITHIN SAM. R

TO

SREE CHITRA TIRUNAL INSTITUTE FOR MEDICAL SCIENCE AND TECHNOLOGY, THIRUVANATHAPURAM

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF

DOCTOR OF PHILISOPHY

2023

DECLARATION

I **NITHIN SAM. R** hereby certify that I had personally carried out the work depicted in the thesis entitled **"GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS"**, except where due acknowledgement has been made in the text.

No part of thesis has been submitted for the award of any degree or diploma prior to this date.

Signature: Nithin Sam

Name of the candidate: **Nithin Sam. R**

Registration number: **2019/PhD/08**

Date: 07-06-2023

Institute for Stem Cell Science and Regenerative Medicine

An autonomous Institutee of
Department of Biotechnology,
Government of India.

inStem

CSCR
CENTER FOR STEM CELL RESEARCH
(a unit of inStem, Bengaluru)
Christian Medical College Campus
Bagayam, Vellore - 632002.

Center for Stem Cell Research

(A unit of inStem, Bengaluru in Collaboration with DBT and CMC, Vellore)

NOT TO BE MINISTERED UNTO BUT TO MINISTER
CHRISTIAN MEDICAL COLLEGE VELLORE INDIA

CERTIFICATE

Name of the Guide: **Dr. Mohankumar K. Murugesan**

Division/Department: **Centre for Stem Cell Research (a Unit of inStem, Bengaluru)**

This is to certificate that **NITHIN SAM. R,** Center for Stem Cell Research of Christian Medical College has fulfilled the requirement prescribed for the Ph.D degree of the Sree Chitra Tirunal Institute for Medical Science and Technology, Thiruvananthapuram.

The thesis entitled, **"GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS"**, was carried out under my direct supervision. No part of the thesis was submitted for the award of any degree or diploma prior this date.

*Clearance was obtained from the Institutional review board, the institutional bio safety committee, the Institutional committee for stem cell research and therapy and Institutional Animal ethics committee for carrying out the study.

Signature of the guide
Dr. Mohankumar K. Murugesan
Scientist E,
Centre for Stem Cell Research
(a Unit of inStem, Bengaluru),
Christian Medical College Campus,

Date: 7/6/23

Christian Medical College Campus,
Bagayam, Vellore - 632002. INDIA.

Phone : +91 - 416 - 228 5101 / 5107
Fax : +91 - 416 - 228 5103
E-mail : cscr@cmcvellore.ac.in

Approval of thesis

The thesis entitled

GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS

Submitted by

NITHIN SAM. R

For the degree of

Doctor of Philosophy

Of

Sree Chitra Tirunal Institute

For

Medical Science and Technology, Thiruvananthapurum

Is evaluated and approved by

Guide: Dr. Mohankumar K. Murugesan

Signature:

Thesis examiner: Dr. Suresh Rayala

Signature:

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and acknowledge the following individuals and institutions who have played a significant role in the completion of my research and thesis:

First and foremost, I would like to extend my sincerest appreciation **to Dr. Mohankumar K. Murugesan**, my research supervisor. I am immensely grateful for his unwavering support, guidance, and provision of both space and funding, which were crucial for the successful completion of my thesis. His invaluable suggestions and constant encouragement have been instrumental in shaping my research journey.

I am grateful to the members of the Dissertation Advisory Committee (DAC): **Dr. Shaji R.V, Dr. Sukesh C. Nair,** and **Dr. Alok Srivastav**, for their insightful suggestions, constructive feedback, and invaluable contributions that have greatly improved the quality of my work.I am grateful to Dr. Shaji R.V, my co-guide from the Department of Hematology, for providing valuable guidance throughout my research. Dr. Sukesh, from the Department of Transfusion Medicine, offered valuable perspectives on coagulation and a deep understanding of megakaryopoiesis. I would also like to express my appreciation to Dr. Alok Srivastav, the Professor and Head of the Centre for Stem Cell Research (CSCR), for providing invaluable suggestions, insights, and access to the state-of-the-art facilities.

I extend my heartfelt thanks to **Dr. Srujan Marepally, Dr. Saravanabhavan Thangavel, Dr. Sanjay Kumar,** and **Dr. Gurbind Singh** for their invaluable contributions as Scientists. Their willingness to invest time, share expertise, and provide essential resources have greatly benefited my research. I am especially grateful to Dr. Srujan for generously providing mRNA for base editing, Dr. Sanjay for supplying AAV

vectors and sharing valuable insights on AAV virus production, and Dr. Gurbind for engaging in thought-provoking discussions that have broadened my perspectives.

I express my sincere gratitude to the collaborators from the Merlin lab, **Dr. Merlin Crossley** and **Henry William Bell**, at the University of New South Wales, Australia. Their assistance with EMSA and ChIP techniques, along with their valuable insights into transcription factors regulating HbF, have been indispensable to my research. I would also like to extend my thanks to **Dr. Jacob E Corn** from the Institute of Molecular Health Sciences, Department of Biology, Switzerland, and **Dr. Stacia K Wyman** from the Innovative Genomics Institute, University of California, Berkeley. Their support in high-throughput sequence analysis and guidance in bioinformatics have been immensely valuable. Furthermore, I am grateful to **Dr. Beeke Wienert**, a former member of the Corn Lab, for her invaluable suggestions and corrections during manuscript preparation. Despite her busy schedule, she patiently explained concepts and consistently responded to my queries, for which I am sincerely appreciative.

I am grateful to my fellow lab mates, **Bhanu, Vignesh, Nivedita, Anila, Kirti, Lokesh,** and **Joshua**, for fostering a supportive and conducive atmosphere within the lab. Their presence has taught me invaluable life lessons, and they have shown patience and understanding towards my mistakes. I am especially appreciative of their care and support when I was unwell. I would like to extend special thanks to Bhanu for offering unique perspectives that have expanded my thinking, and to Anila for her valuable insights into scientific writing and engaging discussions. Their contributions have greatly enriched my research experience.

I would like to express my heartfelt appreciation to my batchmates, **Saranya S, Vignesh R, Vigneshwaran V, Rashme,** and **Aishwarya**, for being a part of my journey. Your companionship has provided moments of relaxation and enjoyment

amidst the challenges. I extend a special thank you to Saranya for engaging in meaningful scientific discussions and guiding me on how to stay updated with research groups. Your insights have been invaluable. I am also grateful to all my friends, including seniors, juniors, and project students from various labs in CSCR. Your diverse knowledge and introduction to new ideas have broadened my horizons and enriched my understanding of different fields.

I would like to express my gratitude to **Mr. Tamil and Mrs. Anupama** for their invaluable assistance and support with office-related tasks during my Ph.D. Their contributions have been greatly appreciated and have played a significant role in facilitating the smooth progress of my research.

I am grateful for the contributions of **Dr. Vigneshwar, Mr. Ashok, Mrs. Pavithra,** and **Mrs. Ester** in organizing the animal facility and their invaluable assistance in conducting animal experiments. Their support has been instrumental in ensuring the smooth operation of the facility and the successful execution of my research involving animal studies.

Special thanks I am grateful to **Mrs. Saranya, Abdul, Immani,** and **Joel** for their support in flow cytometry and organizing core reagents. Special thanks go to **Dr. Sandhya** for her guidance in confocal microscopy, as well as **Mr. Rajesh** and **Mr. Muthu** for their assistance in core facility instrumentation.

To my parents, Dad and Mom **Mr. Ravi Thambi Devadhas** and **Mrs. Sujeya Gracelin**, I apologize for not being able to give you the time that a son should. I am incredibly grateful for your patience, endless support, and unwavering love throughout this journey.

I would also like to acknowledge the funding received from **CSIR** (Council of Scientific and Industrial Research) and the core fund of **CSCR**, which have provided essential financial support for my research.

Lastly, and above all I am incredibly grateful to **God** for showering me with blessings, guiding me through both the highs and lows of life, and molding me into the person I am today. God's presence and guidance have been instrumental in shaping my character and helping me overcome challenges.

"Be still and know that I am God" - Psalm 46:10

Thank you all for your invaluable contributions, guidance, and support. My journey would not have been possible without each and every one of you.

NITHIN SAM. R

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Sl No	Abbreviation	Full Form
1	CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
2	Cas9	CRISPR-associated protein 9
3	dCas9	Dead Cas9
4	nCas9	Nickase CAS9
5	BE	Base Editor
6	ABE	Adenine Base Editor
7	CBE	Cytosine Base Editor
8	SCD	Sickle Cell Disease
9	FIX	Factor 9
10	FVIII	Factor 8
11	FXI	Factor 11
12	HBG	Hemoglobin Subunit Gamma
13	HBA	Hemoglobin Subunit Alpha
14	HBB	Hemoglobin Subunit Beta
15	HBD	Hemoglobin Subunit Delta
16	HBZ	Hemoglobin Subunit Zeta
17	HBE	Hemoglobin Subunit Epsilon
18	HbF	Fetal Hemoglobin
19	HbA	Adult Hemoglobin
20	HbS	Sickle Hemoglobin
21	HbE	Hemoglobin E
22	HbC	Hemoglobin C
23	HbSS	Sickle Hemoglobin homozygous
24	HPFH	Hereditary Persistence of Fetal Hemoglobin
25	ITGA2B	Integrin Subunit Alpha 2B
26	HSC	Hematopoietic Stem Cells
27	HSPCs	Hematopoietic Stem and Progenitor Cells

28	HUDEP-2	Human Umbilical Cord Blood-Derived Erythroid Progenitor
29	HEL	Human Erythroleukemia
30	HEK	Human Embryonic Kidney
31	COS 7	CV-1 in Origin, SV40-transformed
32	PBMNCs	Peripheral Blood Mononuclear Cells
33	CDS	Coding Sequence
34	DNA	Deoxyribonucleic Acid
35	RNA	Ribonucleic Acid
36	cDNA	Complementary DNA
37	DSB	Double Strand Break
38	SSB	Single Strand Break
39	crRNA	CRISPR RNA
40	sgRNA	Single guide RNA
41	gRNA	Guide RNA
42	PAM	Protospacer Adjacent Motif
43	UGI	Uracil DNA Glycosylase Inhibitor
44	spCas9	Streptococcus pyogenes Cas9
45	Indel	Insertion deletion
46	Tad A	tRNA-specific adenosine deaminase
47	APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide
48	PCR	Polymerase Chain Reaction
49	qRT PCR	quantitative Reverse Transcription PCR
50	WLP	Whole Length PCR
51	IOP	In Out PCR
52	mRNA	Messenger RNA
53	HPLC	High-Performance Liquid Chromatography
54	RP-HPLC	Reverse phase-High-Performance Liquid Chromatography
55	ddPCR	Digital droplet PCR

56	FACS	Fluorescence-Activated Cell Sorting
57	HDR	Homologous Directed Repair
58	NHEJ	Non-Homologous End Joining
59	MMEJ	Microhomology-Mediated End Joining
60	GFP	Green Fluorescence Protein
61	SA	Splice Acceptor
62	P2A	porcine teschovirus-1 2A
63	IRES	Internal Ribosome Entry Site
64	EF1α	Elongation Factor 1 Alpha
65	NBSGW	Nonirradiated NOD B6 SCID Il2rγ−/−KitW41/W41
66	PCA	Principal Component Analysis
67	BCL11A	B-cell lymphoma/leukemia 11A
68	KLF	Krüppel-Like Factor
69	TAL	T-cell Acute Lymphoblastic leukemia
70	GATA	GATA binding protein
71	ZBTB7A	Zinc finger and BTB domain-containing protein 7A
72	LCR	Locus Control Region
73	AAV	Adeno Associated Virus
74	ssRNA	Single strand RNA
75	ssODNA	single-stranded Oligonucleotide DNA
76	LV	Lenti-Virus
77	ZFN	Zinc Finger Nuclease
78	TALEN	Transcription Activator-Like Effector Nuclease
79	HA	Hemophilia A
80	HB	Hemophilia B
81	KO	Knock-Out
82	KI	Knock-In
83	Exo	Exonuclease
84	HITI	Homology-Independent Targeted Integration

85	SATI	Single homology Arm donor mediated intron-Targeting Integration
86	HLA	Human Leukocyte Antigen
87	NGS	Next Generation Sequencing
88	RFP	Red Fluorescence Protein
89	SCF	Stem Cell Factor
90	EPO	Erythropoietin
91	TPO	Thrombopoietin
92	FLT3	Fms-like Tyrosine Kinase 3
93	IL	Interleukin
94	IVT	In Vitro Transcription
95	FBS	Fetal Bovine Serum
96	IMDM	Iscove's Modified Dulbecco's Medium
97	RPMI	Roswell Park Memorial Institute
98	DMEM	Dulbecco's Modified Eagle's Medium
99	Opti MEM	Opti-Minimal Essential Medium
100	PS	Penicillin/Streptomycin
101	PBS	Phosphate-Buffered Saline
102	BPBS	BSA-Phosphate-Buffered Saline
103	BSA	Bovine Serum Albumin
104	Edit R	Edit Deconvolution by Inference of Traces in R
105	IGV	Integrative Genomics Viewer
106	COSMID	CRISPR Search with Mismatches, Insertions and/or Deletions
107	IGEAK	Interactive Gene Expression Analysis Kit
108	ICE	Inference of CRISPR Edits
109	GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
110	VEGF	Vascular Endothelial Growth Factor
111	WPRE	Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element
112	APC	Allophycocyanin

113	PE	Phycoerythrin
114	FITC	Fluorescein Isothiocyanate
115	BV	Brilliant Violet
116	DAPI	4',6-Diamidino-2-Phenylindole
117	EMSA	Electrophoretic Mobility Shift Assay
118	ChIP	Chromatin Immunoprecipitation
119	EDTA	Ethylenediaminetetraacetic Acid
120	TBE	Tris-Borate-EDTA
121	IgG	Immunoglobulin G
122	LB	Luria-Bertani
123	OT	Off-Target
124	CD	Cluster of Differentiation
125	hCD	Human CD
126	mCD	Mice CD
127	RBC	Red Blood Cells
128	ANOVA	Analysis of Variance
129	MOI	Multiplicity of Infection
130	VCN	Vector Copy Number

Ph.D SYNOPSIS
GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS

Background:

Hereditary hematological disorders are caused due to faulty genes that are inherited, affecting the cells of hematopoietic lineage. Two major hematological disorders that are common in India are beta hemoglobinopathies and hemophilia. β-hemoglobinopathies are inherited blood disorders caused by pathogenic variants in the beta-globin gene (*HBB*) leading to aberrant or reduced adult-hemoglobin, the protein that carries oxygen in the blood. The most frequent β-hemoglobinopathies are beta-thalassemia and sickle cell disease. In contrast, hemophilia is an inherited bleeding disorders resulting from lack of coagulation factors in the blood, leading to impaired hemostasis. This can lead to spontaneous bleeding as well as bleeding following injuries or surgery. Hemophilia can manifest as either a deficiency in factor VIII or factor IX. Both β-hemoglobinopathies and Hemophilia incur heavy treatment burden due to the need for lifelong medication. The preferred treatment approach for hemoglobinopathies is mostly symptomatic, while for hemophilia is the supplementation of recombinant clotting factors. Although stem cell transplantation from healthy donors can ameliorate beta hemoglobinopathies, it is often limited by the lack of availability of suitable HLA matched donors. Hence, there is a need to develop curative therapeutic approaches for both these disorders.

Genome editing encompasses a range of techniques that enable precise manipulation of the DNA sequence, including insertion, deletion, modification, or replacement of specific genetic elements. This innovative approach holds tremendous

potential for treating genetic disorders. Hematopoietic stem and progenitor cells (HSPCs) editing represents an ideal platform for developing gene editing strategies for hematological disorders as they can be isolated from the patients, manipulated ex vivo, and transplanted back to the patients, thus avoiding the need for matched donors and the risk for GVHD.

In this study, we aimed to optimize two different gene editing strategies, utilizing the revolutionary CRISPR/Cas9 technology in HSPCs, for the treatment of beta hemoglobinopathies and hemophilia B. The initial approach focuses on tackling β-hemoglobinopathies through the utilization of base editing, a technique that enables accurate alterations of individual DNA bases within the targeted region, without inducing double-strand breaks. We hypothesized that creating point mutations in the gamma-globin promoter can mimic Hereditary Persistence of Fetal Hemoglobin (HPFH), a benign genetic condition where the fetal hemoglobin levels in blood is elevated, thereby ameliorating the symptoms caused by the lack of functional beta hemoglobin. We used two base editors namely, Adenine base editor (ABE) that can convert A>G and Cytosine base editor (CBE) that can convert C>T to mutagenize gamma-globin promoter causing disruption of repressor binding sites and creation of activator binding sites resulting in elevated fetal hemoglobin. The second strategy involves the utilization of Homologous Directed Repair (HDR) to facilitate the precise integration of a functional copy of the Factor IX gene into the genome, with the aim of improving the condition of individuals with hemophilia-B. It was hypothesized that the insertion of a single copy of FIX CDS in the platelet-specific promoter, *ITGA2B*, can enable megakaryocytes to produce engineered platelets packed with FIX. These engineered platelets can carry Factor IX and release only at the target bleeding site, thus reducing its exposure to generate and evade inhibitory antibodies during circulation.

Aim:

Optimization of genome editing strategies for the treatment of hereditary hematological disorders

Objectives:

1. To standardize the genome editing tools in the relevant human cellular models
    I. To introduce an array of point mutations into the *HBG* promoter through base editor and then screen for the mutations that induce HbF to therapeutic levels in the human erythroid cell line.
    II. To utilize CRISPR/Cas9 mediated homology directed repair (HDR) strategy for monoallelic integration of human *GFP/FIX* transgene under the platelet specific endogenous promoter (ITGA2B) in HEL (human erythroleukemia) cell line.
2. To validate the editing tools in HSPCs and evaluate their functional outcome
    I. Optimization and validation of base editing technique targeting the gamma globin promoter to induce the expression of fetal hemoglobin in HSPCs.
    II. Optimizing monoallelic integration of *FIX* and *GFP* transgene in *ITGA2B* gene and validating their concerted expression in HSPCs.
3. Elucidating potential novel mechanism, if present, and conducting safety profiling for the edited cells

Experimental methods used in the study:

Two studies were carried out to achieve the proposed objectives.

Study 1: Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin.

Goals:

1. Screening HBG promoter region using ABE and CBE in HUDEP-2 cells (human umbilical cord blood derived erythroid progenitor) to identify novel targets which elevate HbF.
2. Validation of the top target identified from the screening in HUDEP-2 cells.
3. Validation of the best novel targets in CD34+ HSPCs for the evaluation of HbF elevation in the erythroid lineage.
4. Identification of mechanism involved in elevation of HbF in the novel targets.
5. Off-target analysis.

Methodology:

Guide RNAs (gRNAs) targeting *HBG* proximal promoter region spanning up to 300 bp were designed. Arrayed screening of these gRNAs based on lenti-viral approach was carried out in HUDEP-2 stables expressing base editors. The editing efficiency of the gRNAs was evaluated by NGS and Sanger sequencing. The functional outcome of the base edited cells was evaluated by intracellular staining followed by flow cytometry. Among the 42 gRNAs screened, we chose the top 8 gRNAs that gave the highest elevation of fetal hemoglobin using both ABE and CBE and validated further in HUDEP-2 cells using multiple assays. Total fetal hemoglobin positive cells and *HBG* expression before and after erythroid differentiation was assayed using flow cytometry and qRT PCR respectively. The increase in the gamma globin chain was estimated by RP-HPLC, and the total fetal hemoglobin was evaluated by HPLC variants. The erythroid differentiation potential of the cells was evaluated by using flow cytometry. From the validated gRNAs, the novel gRNA that we identified in the screening was further evaluated for its potential to elevate fetal hemoglobin in HSPCs. Adenosine base

editor was delivered as mRNA along with synthetic gRNA. Editing efficiency was evaluated using Sanger sequencing. HbF was evaluated using flow cytometry, qRT PCR, and HPLC. Erythroid differentiation potential and enucleation profile of the cells were evaluated using flow cytometry.

To elucidate the mechanism by which the novel gRNA elevates HbF, we performed EMSA and CHIP qPCR. The DNA off-targets of the guide were evaluated by in-silico prediction followed by targeted deep sequencing. The RNA off-targets were evaluated using transcriptome-wide RNA sequencing. We also performed in vivo study using NBSGW mice to assess the engraftment potential of the edited cells.

Study 2: Monoallelic integration of Factor IX transgene in ITGA2B gene and their concerted expression can ameliorate haemophilia B

Goals:

1. Identification of potential knock-in site in *ITGA2B* locus and standardization of HDR in HEL cells using AAV based *GFP* donor template.
2. Optimization of HDR in CD34+ HSPCs and validating the expression of knock-in genes (*GFP or FIX*) under the endogenous promoter (*ITGA2B*).
3. Characterization of edited CD34+ HSPCs for megakaryocyte differentiation and platelet production.
4. Functional validation of FIX expressed by platelets obtained from the edited HSPCs.
5. Off-target analysis.

Methodology:

The potential site for inserting the *FIX/GFP* CDS was identified in the *ITGA2B* locus (Intron-1) by performing an arrayed screening of gRNAs using lentiviral

approach. The HDR donor construct was prepared by cloning the *GFP/FIX* CDS along with right and left homology flanks in an Adeno Associated Virus 6 (AAV6) backbone. Cas9/gRNA RNP complex was delivered to the cells by electroporation to make the double-strand break, and the donor template was delivered as AAV in HEL cell lines, which are erythroleukemia cell lines where *ITGA2B* is actively expressed. The target gene integration was qualitatively assessed by whole-length PCR (WLP) and In-out PCR (IOP), while the HDR efficiency was quantified using GFP expression by flow cytometry and ddPCR.

The selected target site and the optimized conditions were used as the starting point to standardize HDR in HSPCs similarly. In vitro differentiation of the edited cells into platelet-producing megakaryocytes was achieved. The megakaryopoiesis was estimated by staining with CD41a and CD42b markers via flow cytometry, and the polyploidy nature of megakaryocytes was validated by flow cytometry analysis and imaging using confocal microscopy. The platelets budded from the megakaryocytes were gated based of size, granularity and by staining with CD41a and CD42a markers. Once the *GFP* expression in the platelets was quantified by flow cytometry, the platelets were activated by the agonist TRAP (Thrombin Receptor Activating Peptide). The activated platelets were quantified by CD62p antibody markers using flow cytometry. The FIX expressed in the edited samples were estimated by qRT PCR.

Major finding:

Study I:

1. Lentiviral-based arrayed screening proved to be an optimal method for screening cell lines during base editing experiments.

2. In comparison to Cas9 endonuclease, base editing exhibits minimal or negligible indel formation.
3. Comparative analysis reveals the superiority of base editing over Cas9 when targeting highly homologous regions, such as the HBG promoter encompassing HBG1 and HBG2. Remarkably, base editing demonstrates a significant reduction in the size of the large 4.9 kb deletion.
4. Nickase variant of Cas9 in base editing can occasionally result in low-level double-strand breaks, particularly when targeting highly homologous regions such as the HBG promoter region. These breaks, which occur during replication, can contribute to the occurrence of large deletions.
5. Novel HPFH like point mutations were identified in *HBG* proximal promoter region.
6. The ABE 8e variant emerges as a superior choice compared to ABE 7.10, owing to its hyperactivity. This enhanced activity enables improved editing outcomes and a reduced occurrence of large deletions, attributed to the shorter duration of gRNA binding to the edited sample.
7. Base editing in HSPCs was established; the novel target gRNA-11 (-123 cluster) identified in the study after editing with ABE 8e mRNA gave about 48% fetal hemoglobin compared to 34% by gRNA-2, which targets the known BCL11A repressor binding site (-115 cluster).
8. The mechanism behind the elevation of HbF by editing the novel target was identified. The creation of de novo KLF1 binding site led to the activation of *HBG* expression.
9. Following engraftment in the NBSGW mice model, the editing efficiency of gRNA 11 edited cells remained relatively stable without significant decrease.

Remarkably, a high engraftment rate exceeding 90% (hCD45) was achieved, with no compromise in multilineage differentiation potential.

10. There are no observed Cas-dependent DNA off-target effects or Cas-independent RNA off-target effects associated with the novel target gRNA 11.
11. In this study, a total of five novel clusters were discovered, each of which exhibited an increase in HBG expression. Notably, the -123 cluster demonstrated a comparable elevation of HbF levels to that observed with well-established HPFH mutations.

Study II:

1. The site for target integration in the *ITGA2B* locus was identified. gRNA-10 which gives above 90% indel efficiency and a high MMEJ score was selected in the Intron-1 of *ITGA2B*.
2. The construction of AAV donor plasmids for HDR was optimized. Donors with promoters (Ef1α) and without promoters (SA-P2A and IRES) carrying GFP transgene were constructed.
3. Strategies for HDR were established in the HEL cell lines using GFP based donors. Up to 80% HDR efficiency was obtained in HEL cells.
4. Optimization of efficient HDR strategies was achieved for both GFP and FIX donors in HSPCs, resulting in up to 40% target integration. Additionally, successful demonstration of efficient megakaryopoiesis of the edited cells was observed in vitro.
5. It was demonstrated that monoallelic integration of the GFP gene in *ITGA2B* intron could express GFP without compromising *ITGA2B* expression.
6. Platelets derived from the edited samples exhibited GFP packaging and co-expression of CD41a (ITGA2B).

7. Cas-dependent DNA off-target analysis was performed and no off-targets were identified.

Significance of the finding:

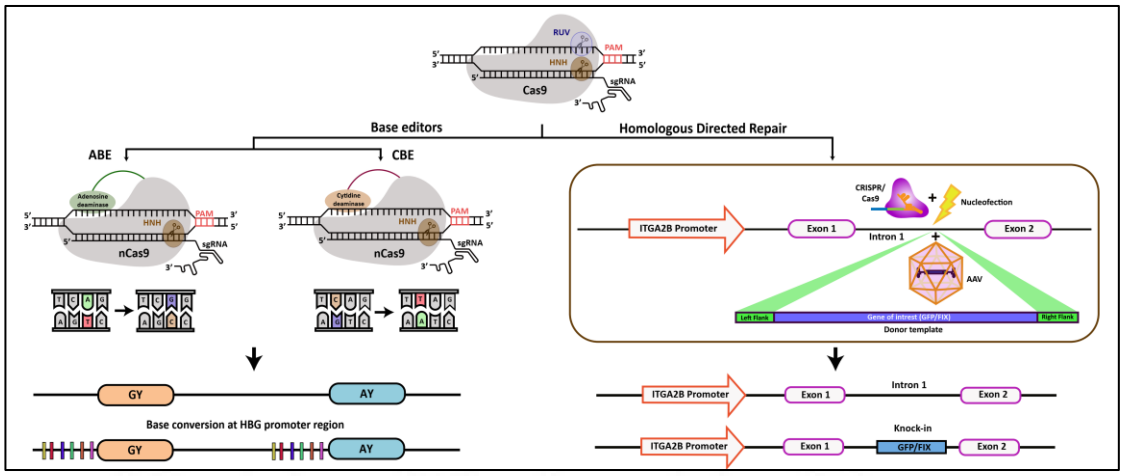
In the first study:

1. Provided evidence that base editing reduces large deletions when highly homologous regions like *HBG* proximal promoter.
2. Through base editing screening, novel regulator element binding sites can be identified.
3. The study identified targets that effectively enhance fetal hemoglobin levels, showcasing their potential for therapeutic gene editing in beta hemoglobinopathies.
4. The novel target identified was patented and will be taken to preclinical studies to confirm the translational potential.

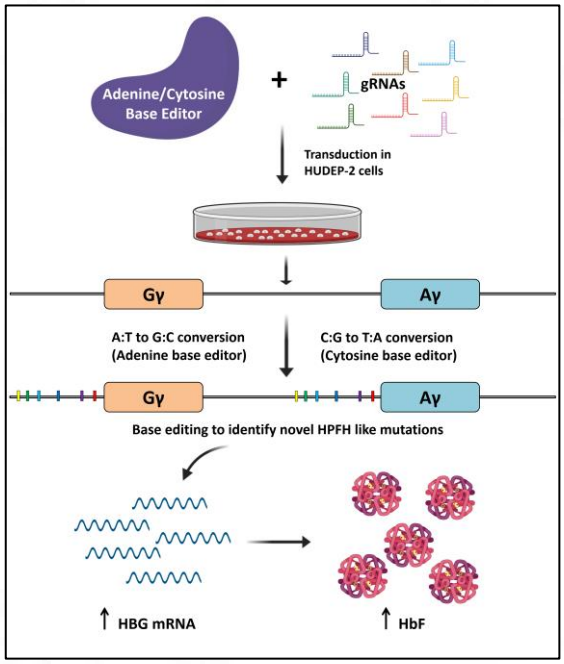
In the second study:

1. Concerted expression by monoallelic integration of *GFP/FIX* transgene in the *ITGA2B* gene was established in cellular model and HSPCs.
2. In this study, an HDR-based approach was employed to integrate the *GFP/FIX* transgene under the platelet-specific promoter, mitigating the risk of random integration.
3. While the successful integration of the *FIX* transgene in HSPCs was achieved, quantification of FIX was not feasible due to the inability to obtain platelets from in vitro culture. Further validation in an ex vivo mice model is necessary to quantify the expression of FIX.

Graphical overview of the thesis

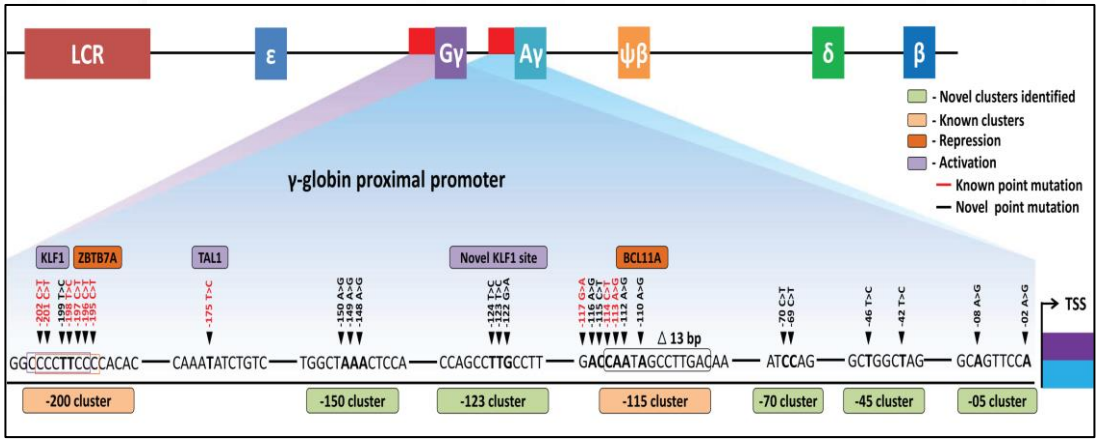


Study 1: Graphical overview

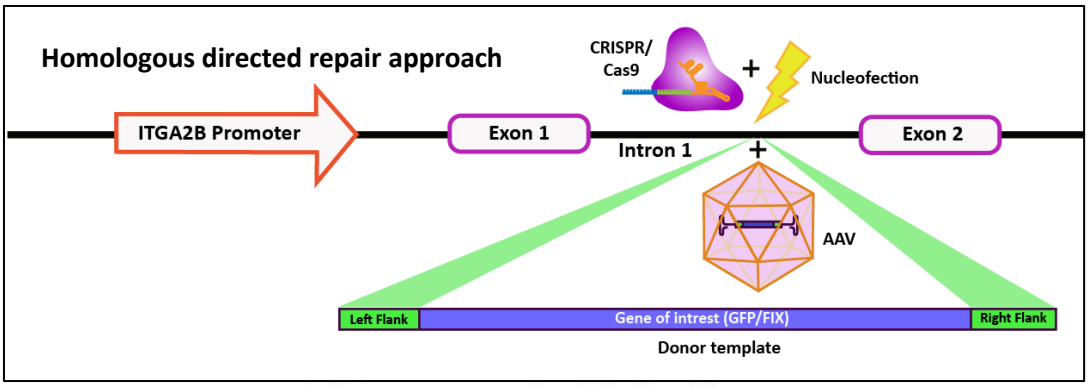


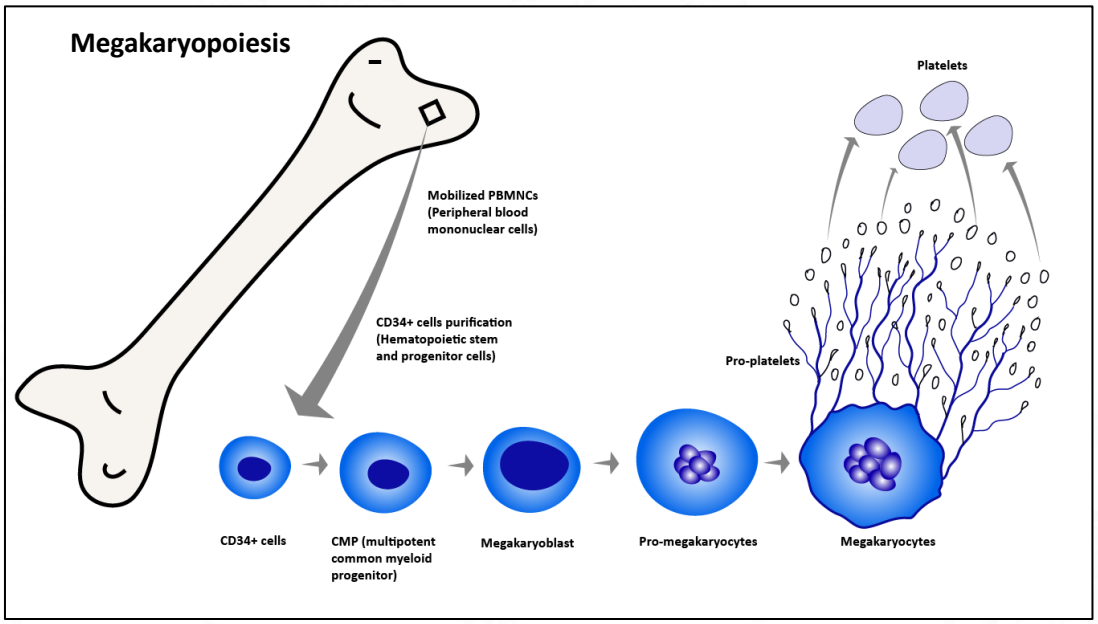
Creation of HPFH like point mutation to elevate Fetal hemoglobin

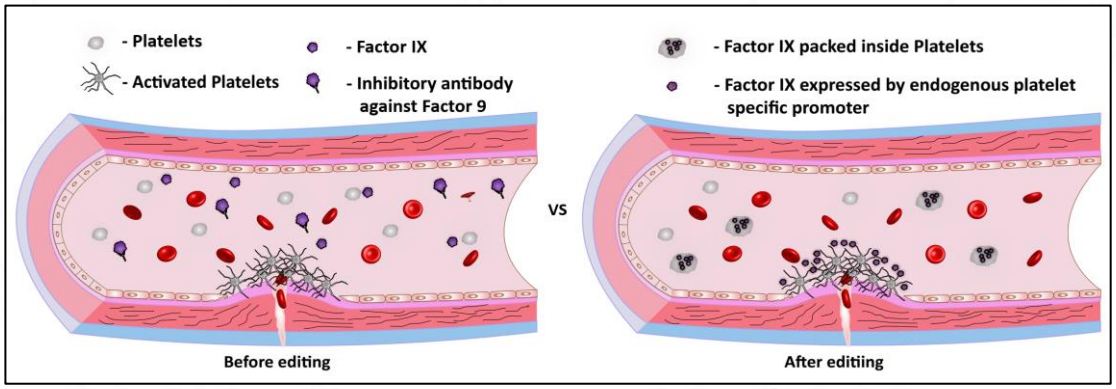
Novel clusters identified in HBG promoter region which elevates HbF



Study 2: Graphical overview







Edited cells were subjected to megakaryopoiesis to obtain engineered platelet packed with FIX/GFP.

INTRODUCTION

1.1 Background of the study

Hereditary hematological disorders are a group of inherited disorders that affect the composition or function of blood. Apart from hematological malignancies the most common genetic blood disorders are hemoglobinopathies and hemophilia. Hemoglobinopathies are caused by mutations that affect the structure or quantity of hemoglobin, while hemophilia is caused by the deficiency of proteins in the coagulation cascade. To date there is no curative therapy for hemoglobinopathies and hemophilia, except bone marrow transplantation from healthy matched donor, which is the only curative treatment available. Hence both the diseases entail lifelong treatment burden and reduces the quality of life. Gene therapy is an attractive treatment strategy that promises one shot cure by correcting or bypassing the underlying genetic mutation (Al-Saif, 2019).

Beta-hemoglobinopathies are autosomal recessive diseases that affect the beta globin gene, the predominant disorders are β -thalassemia and sickle cell disease (SCD). β -thalassemia is caused by a spectrum of mutations in the beta globin gene that usually affects its synthesis or produces abnormal beta globin chains that are non- functional. SCD on the other hand is caused by a single point mutation in the 6^{th} codon of beta globin, resulting in an amino acid substitution that causes hemoglobin to precipitate under hypoxia, resulting in hemolytic anemia. Both β-thalassemia and SCD are highly prevalent across the globe and the disease burden is high in places where consanguinity is a common practice. The disease manifestations of both the diseases vary significantly with β-thalassemia mostly affecting the production of hemoglobin and sickle cell anemia affecting the RBC quality. The

treatment for both the diseases is mostly symptomatic and patients often require repeated blood transfusions. The only curative approach available is stem cell transplantation from a healthy matched donor but carries the risk of graft rejection and other immune complications (Barbarani *et al*., 2020; Frati and Miccio, 2021).

Hemophilia is an X linked recessive disorder caused by mutations that affect the production of coagulation factors. The two major types are hemophilia A, caused by deficiency of Factor VIII and hemophilia B, caused by deficiency of Factor IX. The main clinical manifestation of hemophilia is uncontrollable bleeding, either spontaneous or injury related, and can even be life threatening. The approved treatment option for hemophilia is recombinant coagulation factor therapy which has improved the median life expectancy of the patients tremendously but is expensive and limited by the development of inhibitory antibodies (Berntorp *et al*., 2021).

Gene therapy represents a promising strategy for treating genetic disorders through the correction of causative mutations or the delivery of healthy copies of genes. A variety of genome editing tools have been developed, with the CRISPR/Cas9 system standing out for its versatility. Specifically, Cas9 nuclease can make precise double-strand breaks in the genome at target locations, which can then be repaired by error-prone DNA repair pathways in the cells. By providing a donor template with the desired mutations, this repair can be made error-free and integrated into the genome via homology-directed repair, thereby allowing for the insertion of DNA sequences ranging from a few nucleotides to entire genes. Furthermore, the fusion of deaminases to Cas9 nickase can enable the precise conversion of nucleotides without inducing double-strand breaks, a process known as base editing. There are two commonly used base editors, including the Adenine base editor (which converts adenosine to guanosine) and the Cytosine base editor (which converts

cytosine to thymine), which can be used to minimize indels (Adli, 2018; Li *et al*., 2023; Porteus, 2019; Prasad *et al*., 2021; Tröder and Zevnik, 2022).

The objective of this study was to develop gene editing strategies for two common hematological disorders: β-hemoglobinopathies and hemophilia B. For β-hemoglobinopathies, we developed a strategy to reactivate developmentally silenced fetal hemoglobin expression by manipulating the promoter sequence of the gamma-globin gene (HBG). Specifically, we hypothesized that creating point mutations in the gamma-globin promoter could mimic Hereditary Persistence of Fetal Hemoglobin (HPFH), a benign genetic condition that elevates fetal hemoglobin levels in blood and ameliorates symptoms caused by the lack of functional beta hemoglobin (Wienert *et al*., 2018). To achieve this, we used two base editors, the Adenine base editor (ABE) and the Cytosine base editor (CBE), to mutagenize the gamma-globin promoter thus disrupting repressor binding sites and creating activator binding sites, ultimately resulting in elevated fetal hemoglobin. For hemophilia, we hypothesized that the insertion of a single copy of FIX CDS (Padua variant) in the platelet-specific promoter, ITGA2B, could enable megakaryocytes to produce engineered platelets packed with FIX. These engineered platelets can carry Factor IX and are only released at the target bleeding site, reducing exposure to inhibitory antibodies during circulation.

1.2 Specific aim and objectives of the study

The purpose of this research study is to develop and optimize two genome editing strategies for the treatment of hereditary hematological disorders, β-hemoglobinopathies and hemophilia-B, in healthy donor derived Hematopoietic Stem and Progenitor Cells (HSPCs). The first goal is to standardize genome editing

strategies in respective cell lines. HUDEP-2 cell line will be used to optimize base editing, while HEL cell line will be used for knocking in GFP/FIX transgeen under ITGA2B promoter via HDR. The editing and functional assays in both the strategies will be validated using Sanger sequencing, NGS, flow cytometry, PCR, qRT PCR, ddPCR and HPLC.

After standardizing base editing and HDR in respective cell lines, the second objective is to optimize these strategies further in the healthy donor HSPCs. For base editing, HSPCs will be edited by electroporating the synthetic gRNA along with base editor mRNA, followed by the functional validation of HbF in the erythroid cells derived from the HSPCs. For HDR, the RNP complex consisting of synthetic gRNA targeting the specific locus along with the Cas9 will be delivered to the HSPCs via electroporation. Following electroporation, the cells will be transduced with AAV6 carrying the donor template. The cells will be differentiated toward megakaryocytic lineage and the platelets budded from megakaryocytes will be analyzed for GFP/FIX expression.

The third objective of this study is to test the safety of the edited cells. In case of base editing, the edited cells will be analyzed for Cas independent RNA off-target and Cas dependent DNA off-target, while in case of gene knock-in, only Cas dependent DNA off-target will be validated. The edited HSPCs will be differentiated to the specific lineage of intent to analyze any defect in differentiation. The engraftment potential and multilineage differentiation potential of these edited cells will be validated in NBSGW mice model. Overall, this research aims to optimize genome editing strategies for the treatment of hereditary hematological disorders and ensure the safety and efficacy of edited HSPCs.

Key objectives

1. To standardize the genome editing tools in the relevant human cellular models
    I. To introduce an array of point mutations into the *HBG* promoter through base editor and then screen for the mutations that induce HbF to therapeutic levels in the human erythroid cell line.
    II. To utilize CRISPR/Cas9 mediated Homology Directed repair (HDR) strategy for monoallelic integration of human FIX transgene under the platelet specific endogenous promoter (*ITGA2B*) in HEL (human erythroleukemia) cell line.
2. To validate the editing tools in HSPCs and evaluate their functional outcome
    I. Optimization and validation of base editing technique targeting the gamma globin promoter to induce the expression of fetal hemoglobin in HSPCs.
    II. Optimizing monoallelic integration of Factor IX and GFP transgene in *ITGA2B* gene and validating their concerted expression in HSPCs.
3. Elucidating potential novel mechanism, if present, and conducting safety profiling for the edited cells

REVIEW OF LITERATURE

2.1 Hereditary hematological disorders

Blood is a multifaceted, viscous fluid that flows through the vasculature of the body, serving a fundamental function in the transportation of oxygen, nutrients, hormones, and other biologically important molecules to various tissues. Blood serves as a means of removing carbon dioxide, metabolic waste, and superfluous compounds from the body. The composition of blood is heterogeneous, consisting of different components such as red blood cells, white blood cells, platelets, and plasma. Hemoglobin, a protein that binds oxygen and transports it from the lungs to the tissues, is a critical component of red blood cells, while white blood cells are an essential part of the immune system that aids in the prevention of infections. Platelets play a significant role in the prevention of excessive bleeding by forming clots, while plasma is a yellowish fluid that contains various proteins, electrolytes, and other vital substances. An average adult has around 5 liters of blood in their body, and blood cells are generated in the bone marrow, which is the spongy tissue located within certain bones. Blood is crucial in maintaining homeostasis, the internal balance of the body, and is involved in a variety of physiological processes such as respiration, digestion, and temperature regulation.

Hematology is a field of medicine that deals with the scientific study of blood and blood-forming tissues, as well as the diagnosis, treatment, and prevention of blood-related disorders. The scope of hematology includes various conditions such as anemia, clotting disorders, bleeding disorders, blood cancers, and other hematological disorders. Hereditary hematological disorders are a type of blood disorder that occurs due to genetic mutations inherited from parents. These disorders

affect the production, function, or structure of blood cells and components such as red blood cells, white blood cells, platelets, and plasma proteins. The major types of hereditary hematological disorders are coagulation factor deficiencies, including hemophilia A and B and Von Willebrand disease; Red blood cell disorders such as sickle cell disease, thalassemia, hereditary spherocytosis, glucose-6-phosphate dehydrogenase deficiency, hereditary elliptocytosis and congenital dyserythropoietic anemia; and bone marrow disorders such as Fanconi anemia and Diamond-Blackfan anemia. Some of these disorders have high prevalence and are classified as major hereditary hematological disorders, including Beta hemoglobinopathies and hemophilia (Al-Saif, 2019; Barbarani *et al*., 2020; Berntorp *et al*., 2021; Frati and Miccio, 2021).

2.2 Beta hemoglobinopathies

The human alpha- and beta-globin loci, situated on chromosomes 16 and 11 respectively, exhibit a sequential pattern of gene activation and silencing during development. This process gives rise to various forms of hemoglobin, including embryonic, fetal, and adult types, which are expressed at distinct stages. In early embryonic development, the α-globin locus undergoes the expression of ζ-globin (HBZ), followed by its subsequent silencing in the first trimester. Subsequently, α-globin (HBA1/HBA2) expression increases and persists throughout adulthood. The β-globin locus displays a more intricate hemoglobin switching mechanism, involving two transitions. The first transition occurs simultaneously with the ζ- to α-globin switch, while the second occurs around the time of birth. During the first trimester of pregnancy, ε-globin (HBE) predominates as the primary β-like globin chain, forming functional embryonic hemoglobin (Hb Gower-1 or Hb Gower-2) in combination with

ζ-globin or α-globin. As ε-globin expression diminishes, duplicated fetal γ-globin genes (A gamma or HBG1 and G gamma or HBG2) become upregulated. The association of γ-globin polypeptides with α-globin chains gives rise to the principal fetal hemoglobin, hemoglobin F (HbF). The second transition occurs shortly after birth, when γ-globin expression is largely silenced, and adult β-globin (HBB) assumes the role of the major β-like globin. Additionally, a small amount of δ-globin (HBD) is expressed in adulthood. The combination of β- and δ-globin chains with α-globin results in the formation of adult hemoglobin A (HbA) and adult hemoglobin A2 (HbA2), respectively. At the molecular level, the process of hemoglobin switching involves the establishment of long-range chromatin interactions between a common locus control region (LCR) and specific globin promoters active during distinct developmental stages. This molecular orchestration takes place within an active chromatin hub (ACH) and relies on the presence of transcription factors or cofactors that bind to specific DNA consensus sequences, facilitating favorable conditions for gene expression.(Wienert *et al*., 2018).

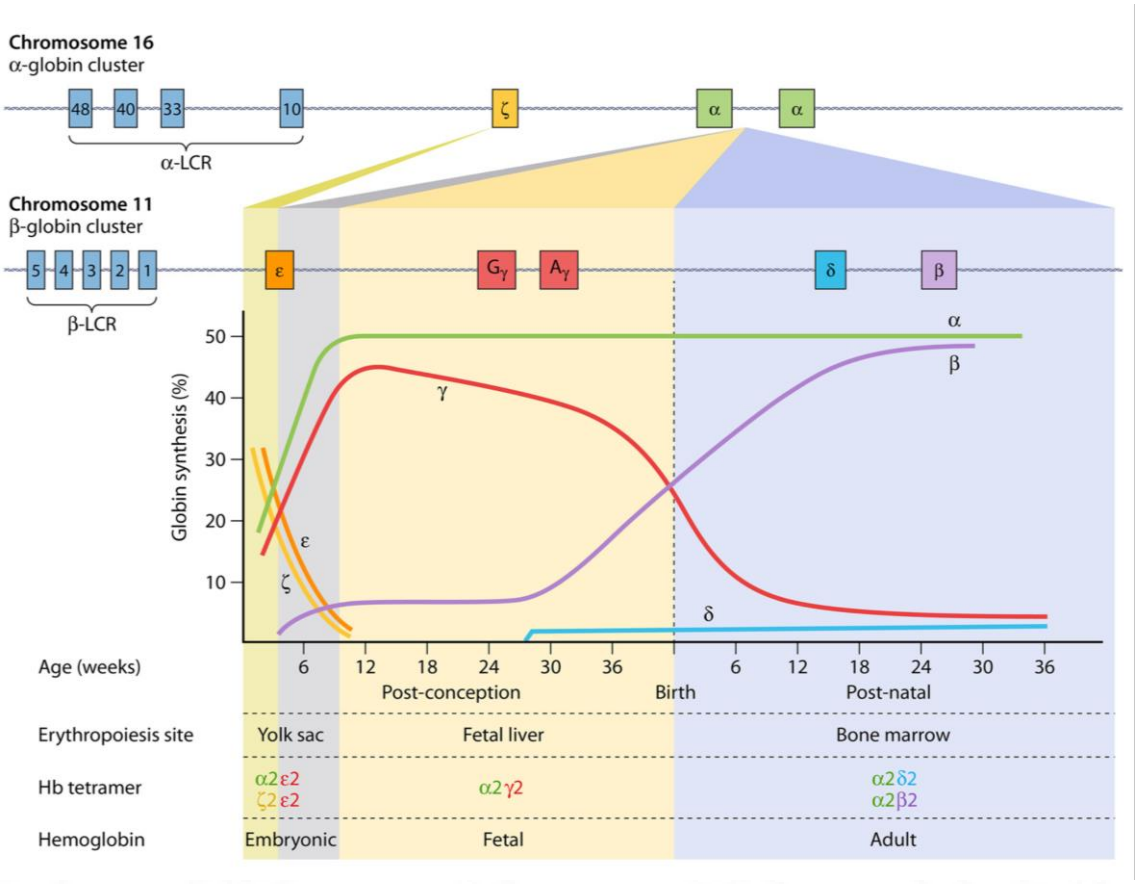


Figure 2.1: Temporal Dynamics of Globin Gene Regulation during Human Development: From Embryonic to Fetal to Adult Hemoglobin
(Source - Venkatesan V *et al*., Mol Cell Biol. 2020 Dec 21;41).

Beta hemoglobinopathies encompass a group of genetic disorders affecting red blood cells, ranging from mild to severe anemia. These disorders arise from abnormalities in the β-globin chains of hemoglobin, which is responsible for oxygen transport in the blood. Structural alterations in hemoglobin or reduced/absent hemoglobin production, known as β-thalassemia, contribute to the manifestation of symptoms. Beta thalassemia can be categorized as major or intermedia, with major cases typically presenting symptoms before the age of two, while symptoms in intermedia cases appear later in life. Examples of beta hemoglobinopathies caused by structural changes in hemoglobin include HbS (sickle cell disease), HbE, and HbC. Hematopoietic stem cell transplantation is currently the definitive cure for β-

hemoglobinopathies, although it is only available to a limited number of patients who have an HLA-matched donor. The effectiveness of treating β-thalassemia and sickle cell disease (SCD) is currently limited, with hydroxyurea being the only medication that demonstrates some efficacy. However, advancements in genome modifications offer promising avenues for addressing β-hemoglobinopathies.

2.2.1 Sickle Cell Disease:

The substitution of β6Glu>Val amino acid in SCD leads to the formation of long hydrophobic polymers of HbS, which precipitate within the cell under hypoxic conditions, resulting in the typical sickle shape. This sickling phenomenon causes vessel obstruction due to cell adhesion and also causes hemolysis due to fragility, eventually leading to anemia. Multiple organs, such as lungs, kidneys, and the retina, can be damaged in SCD, which is associated with an average life expectancy of 20 years less than the general population. The severe homozygous form (HbSS disease) affects approximately 300,000 infants worldwide each year, with over half of cases occurring in Nigeria, India, and the Democratic Republic of Congo. SCD is also prevalent in the Mediterranean region and the Middle East and is now widespread worldwide due to migration (Rees *et al*., 2010).

2.2.2 β-thalassemia:

Mutations in the β-globin gene result in β-thalassemia, a range of disorders that can vary in severity from complete absence (β0) to partial reduction (β+). These mutations can be caused by a wide range of genetic alterations, including deletions, insertions, or point mutations, which affect the various steps involved in β-globin gene regulation. This includes transcription, RNA processing, and translation. Ineffective erythropoiesis and hemolysis, leading to reduced production of functional

β chains and anemia. Mutations outside the β-locus, such as those found in *XPD* or *GATA1*, can also lead to β-thalassemia in rare cases. The condition is most prevalent in individuals of Mediterranean, Middle Eastern, and Southeast Asian descent, affecting approximately 1.5% of the global population, with the highest prevalence in countries like Iran, Pakistan, and India (Frangoul *et al*., 2021).

2.3 Hereditary persistence of fetal hemoglobin (HPFH):

In adults, the levels of hemoglobin F (HbF) typically constitute less than 1% of total hemoglobin. However, individuals with hereditary persistence of fetal hemoglobin (HPFH) syndrome exhibit incomplete silencing of fetal globin genes, resulting in higher HbF levels. Even modest increases in HbF can have beneficial effects, and levels above 20% are considered sufficient to prevent clinical crises. The extent of HbF elevation can influence the symptoms associated with beta-hemoglobinopathies. Extensive research spanning nearly 50 years has been dedicated to understanding HPFH, aiming to achieve widespread reactivation of HbF expression for optimal clinical outcomes. HPFH can be classified into two primary categories: deletional and non-deletional HPFH (Wienert *et al*., 2018).

2.3.1 Deletional HPFH:

Deletional HPFH is characterized by substantial deletions in the DNA region spanning between the γ- and β-globin genes within the globin locus. These deletions have been associated with elevated levels of hemoglobin F (HbF) and are linked to at least nine distinct major deletions. The loss of the δ- and β-globin genes due to these deletions gives rise to both δβ-thalassemia and HPFH. In deletional HPFH cases, occasional deletions encompass a critical regulatory element located in close

proximity to the ψβ pseudogene. Through three-dimensional analysis of chromosomal architecture, it has been revealed that this region serves as a vital anchoring element. Within the β-globin locus, it interacts with other elements to establish spatial segregation between fetal and adult globin genes. Additionally, depending on the specific developmental stage, it facilitates long-range interactions with the locus control region (LCR). Notably, BCL11A has been identified as a key regulator of the pseudo beta region, as its knockout leads to similar outcomes as the CRISPR-Cas9 deletion of the ψβ pseudogene. Recent investigations have suggested that BCL11A directly binds to this region, further implicating its role in the regulation of HbF expression (Venkatesan *et al*., 2021).

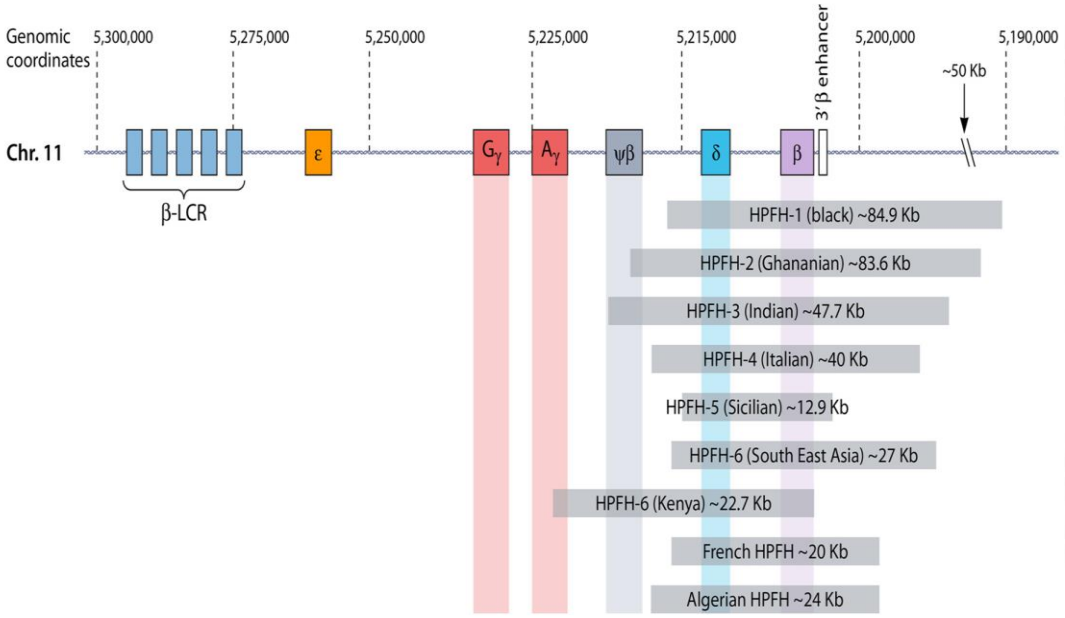


Figure 2.2: Schematic Representation of Large Deletions in the β-Globin Cluster Associated with Deletional Hereditary Persistence of Fetal Hemoglobin (HPFH)
(Source - Venkatesan V *et al*., Mol Cell Biol. 2020 Dec 21;41).

2.3.2 Non-deletional HPFH:

The promoters of proximal fetal γ-globin genes often harbor point mutations or minor deletions that give rise to non-deletional hereditary persistence of fetal

hemoglobin (HPFH). Through evolutionary duplication of the γ-globin genes, two highly similar genes, HBG1 (also known as A-gamma) and HBG2 (commonly referred to as G-gamma), emerged, both of which have been found to harbor HPFH mutations in their promoters. These genetic alterations, encompassing small insertions, deletions, or point mutations, are frequently identified in unaffected family members during genetic screening for hemoglobinopathies, indicating their benign nature. However, in the context of other genetic disorders within the β-globin locus, such as sickle cell disease (SCD) and β-thalassemia, these mutations can have a beneficial effect by reducing symptoms through elevated levels of fetal hemoglobin (HbF) (Wienert *et al*., 2018).

Non-deletional HPFH patients possess point mutations or minor deletions that influence the promoters of proximal γ-globin genes. The proximal promoter region exhibits two clusters of mutations: one located approximately -200 relative to the transcriptional start site and another around -115. Notably, a single point mutation at position -175 stands out as the sole known HPFH mutation in isolation. By disrupting the binding of transcriptional repressors necessary for the normal silencing of γ-globin genes, these mutations generate two densely populated clusters around the -115 and -200 positions, resulting in sustained de-repression of γ-globin and subsequent elevation of HbF levels. While the identity of candidate repressors binding to γ-globin promoters lacked in vivo evidence, the identification of BCL11A and ZBTB7A as crucial in vivo HbF repressors has shed light on their repressive mechanisms. Recent studies involving genetically modified cells carrying HPFH promoter mutations have revealed that ZBTB7A binds to the -200 location of the γ-globin promoter, and this binding is disrupted by HPFH mutations in that region.

Similarly, it has been discovered that the repressor BCL11A binds to the -115 position, and the ability of BCL11A to bind the promoter directly is compromised by HPFH mutations in the -115 area. Consequently, current research efforts are focused on elucidating the direct repression of promoters through well-defined critical elements bound by BCL11A and ZBTB7A (Wienert *et al*., 2018).

However, not all non-deletional HPFH mutations exert their function by altering repressor binding. One such mutation is the -175T>C mutation, which introduces a new binding site for the erythroid activator TAL1. The expression of fetal globin genes is facilitated by TAL1's recruitment of cofactors and the locus control region (LCR) to the γ-globin promoter. Another noteworthy example is the -198T>C mutation, which represents a gain-of-function mutation creating a binding site for the erythroid activator KLF1. Interestingly, this mutation occurs within the cluster of HPFH mutations that impede ZBTB7A binding. These findings suggest that the introduction of novel binding sites for activators can tilt the balance towards increased expression of fetal γ-globin. TAL1 and KLF1, as the primary mediators of globin gene activation, serve as erythroid master regulators involved in the regulation of globin genes. The existence of novel binding sites for these elements within the γ-globin promoter enables similar regulatory mechanisms governing adult γ-globin production (Wienert *et al*., 2018).

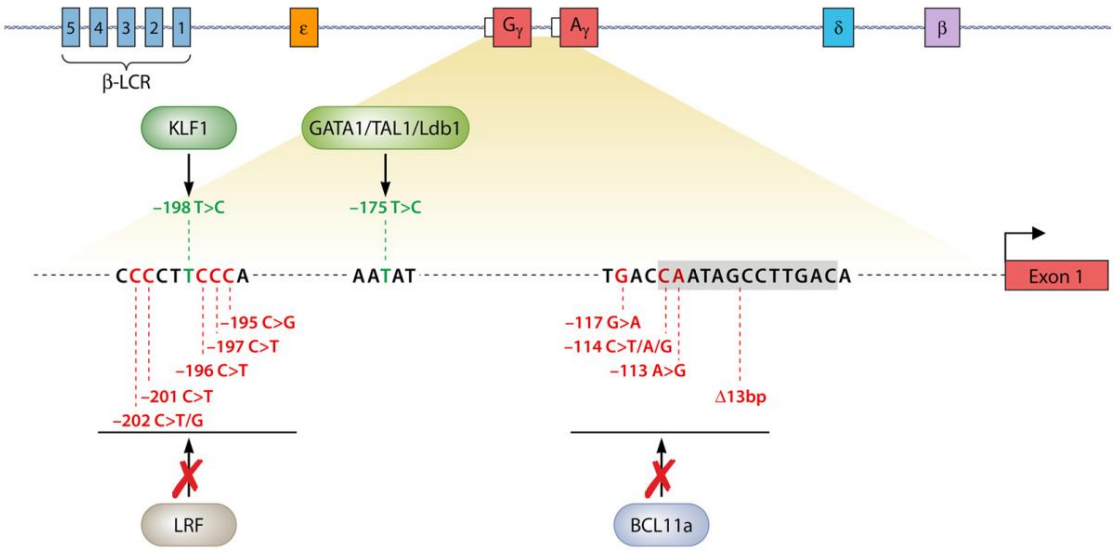


Figure 2.3: Schematic Representation of substitution in the γ-Globin Promoter Associated with Non-deletional Hereditary Persistence of Fetal Hemoglobin (HPFH)
(Source - Venkatesan V *et al*., Mol Cell Biol. 2020 Dec 21;41)**.**

2.4 Hemophilia

Hemophilia A and hemophilia B are genetic disorders resulting from the deficiency or absence of two coagulation proteins - factor VIII (FVIII) (caused by a mutation in the F8 gene) and factor IX (FIX) (caused by a mutation in the F9 gene) respectively. Hemophilia C used to be defined as a deficiency of coagulation factor XI (FXI), but it is no longer considered a form of hemophilia. The severity of hemophilia is determined by the type of genetic mutation, and the degree of severity is classified as severe, moderate, or mild based on the amount of residual FVIII or FIX activity. Hemophilia A and B predominantly affect males due to their X-linked inheritance pattern, while females are usually carriers of one mutated gene and may have mild symptoms. The symptoms of hemophilia are similar in both types and include bleeding into joints, especially large ones such as the elbows, knees, and ankles, which can lead to painful and disabling hemophilic arthropathy. However, life-threatening bleeds may also occur, such as intracranial bleeds and bleeds in other

internal organs, regardless of the severity of the disease, but are more common in severe forms (Berntorp *et al*., 2021).

Determining the prevalence of hemophilia is challenging due to the availability of treatment and differences in life expectancy. Accurate epidemiological data is reliant on adequate diagnosis, registration, and treatment. Hemophilia A is more prevalent than hemophilia B, and both types affect all ethnic groups equally. While the reported prevalence of hemophilia A in the male population is 1 in 5,000 and 1 in 10,000 overall, more precise estimates demonstrate that access to diagnosis, registration, and treatment significantly impact prevalence. A study of 106 countries revealed that the prevalence of hemophilia A in high-income countries was estimated at 12.8 per 100,000 males, whereas low-income countries had a prevalence of 6.6 per 100,000 males. The reported prevalence of hemophilia B in males is 1 in 30,000. A review of data from 105 countries found that the prevalence of hemophilia B in high-income countries was 2.7 per 100,000 males, while in low-income countries, it was 1.2 per 100,000 males (Berntorp *et al*., 2021).

2.4.1 Hemophilia A

Hemophilia A is a bleeding disorder inherited through an X-linked recessive pattern, caused by mutations in the *FVIII* gene, which is made up of 26 exons located on the distal end of the X chromosome long arm (Xq28) and codes for the FVIII protein, consisting of heavy and light chains. The mutations causing hemophilia A are diverse, including large deletions, nonsense mutations, inversion of intron 22 or intron 1, and missense mutations, with inversion of intron 22 being the most frequent cause for severe hemophilia A. Replacement factor therapy with intravenous therapeutic FVIII (tFVIII) is the main treatment for hemophilia A, but the development of inhibitors against tFVIII can complicate the therapy, particularly in

people with severe hemophilia A caused by more disruptive F8 genotypes, where inhibitor risk is up to 40%, often within the first 50 days of tFVIII exposure and typically early in life (Berntorp *et al*., 2021).

2.4.2 Hemophilia B

The gene coding for FIX is substantially a smaller protein consisting of 415 amino acids, and is synthesized in hepatocytes. Unlike FVIII, FIX requires vitamin K for enzymatic carboxylation of the glutamic acid residues of the Gla domain. Defective F9 leads to hemophilia B, an X-linked recessive inherited bleeding disorder. Missense mutations are the most common cause of hemophilia B. The Padua variant (R338L), which is a gain-of-function FIX variant, is responsible for a rare X-linked thrombophilia. It has been used successfully to improve hemophilia B gene therapy strategies. FIX activation involves two processes: FVIIa-dependent activation during the initiation of hemostasis and the activation of FXI (FXIa) by thrombin. The carboxylation of glutamic acid residues triggers a conformational optimization of the Gla domain after the exchange of Mg2+ for Ca2+. This conformational change enables phospholipid binding by the Gla domain. FIXa binds to FVIIIa, enabling hydrolysis of FX to the active form FXa, which is the enzymatic contributor to the downstream prothrombinase complex, generating thrombin (Berntorp *et al*., 2021).

2.5 Viral vector-mediated gene therapy for beta-hemoglobinopathies and hemophilia

The range of viral vectors available for gene therapy is wide and includes both transient and long-term expression. The vectors can be RNA or DNA viruses with single or double-stranded genomes. Adenoviruses are the most commonly used

viral vectors due to their packaging capacity of 7.5 kb of foreign DNA, which provides short-term episomal expression in a broad range of host cells. However, they can generate strong immune responses. AAV vectors, on the other hand, carry a small ssRNA genome, which allows packaging of only 4 kb inserts but provides long-term transgene expression through chromosomal integration with low pathogenicity and toxicity. However, their limited packaging capacity and potential immune response triggered by repeated administration are limitations. Retroviruses are the gold standard vectors for long-term gene therapy but can only infect dividing cells. Lentiviruses, which belong to the retrovirus family, can infect both dividing and non-dividing cells, providing low cytotoxicity, making them attractive for therapeutic applications requiring long-term expression. Lentiviruses possess the same packaging capacity and chromosomal integration as conventional retroviruses (Rashighi and Harris, 2017)(Batty and Lillicrap, 2021).

2.5.1 Viral vector-mediated gene therapy for beta-hemoglobinopathies

The conventional approach for treating β-hemoglobinopathies involves introducing a globin transgene into hematopoietic stem and progenitor cells (HSPCs) of a patient. This transgene can either be an additional copy of an existing globin gene or a modified globin gene that can reduce the severity of the disease. Gene addition therapy utilizes lentiviral vectors (LVVs) to integrate a transgene into a target cell, allowing for stable and durable expression of therapeutic genes. LVVs are considered safe and effective, capable of delivering large DNA payloads and transducing both dividing and nondividing cells. This approach enables the incorporation of specific regulatory elements and provides traceability of transduced cells and their progeny. Multiple clinical trials are underway, focusing on treating diseases such as transfusion-dependent thalassemia (TDT) and sickle cell disease

(SCD) using various gene addition strategies. Promising results have been observed, including significant reductions in transfusion requirements and stable production of therapeutic hemoglobin. (Leonard *et al*., 2022).

Clinical trials using additive globin techniques have been successful in treating β-thalassemia but have yet to be carried out for α-thalassemia. Various research centers and hospitals are sponsoring several ongoing clinical trials for SCD and β-thalassemia. Lentiviral vectors have been widely employed in these trials to transfer the globin transgenes to the HSPCs. Successful correction of β0/βE thalassemia major has been demonstrated using a β-globin vector carrying a T87Q mutation, which resulted in a substantial increase in hemoglobin levels and independence from transfusions. The same vector has also been utilized to correct the sickle phenotype in at least one patient. While cases of acute myeloid leukemia (AML) have been reported in SCD patients following gene addition therapy, it is believed to be unrelated to the lentiviral vectors used. It is important to note that no evidence of genotoxicity has been found in the long-term follow-up of patients treated with lentivirus vectors for hemoglobinopathies and other diseases (Locatelli *et al*., 2022).

2.5.2 Viral vector-mediated gene therapy for hemophilia

Gene therapy is a promising treatment for hemophilia, a bleeding disorder caused by a deficiency in coagulation factors. Adeno-associated viral (AAV) vector-based gene therapy has been successful in preclinical and clinical studies for hemophilia. The limited packaging capacity of AAV vectors led to a focus on hemophilia B (HB) due to the small size of the *FIX* gene. AAV vectors use a codon-optimized transgene under the control of a liver-specific promoter to restrict expression to hepatocytes, and several AAV vector serotypes are being investigated, all of which exhibit liver tropism. The immunogenicity of AAV is the main limitation,

with neutralizing antibodies to the AAV capsid occurring in 30-40% of the general population, and the capsid-mediated cellular immune response resulting in transient hepatotoxicity. The first long-term expression of AAV liver gene therapy used AAV8 and resulted in stable transgene expression levels and improved disease phenotype. The next generation of studies used a FIX variant, FIX-Padua, which has enhanced protein-specific activity, resulting in higher expression levels (Leebeek and Miesbach, 2021).

One of the major challenge in treating HA with gene therapy is related to packaging the vector with the FVIII gene, which is significantly larger (7 kb) than the packaging capacity of AAVs, which is around 5 kb. Factor VIII consists of three A domains, one B domain (about 44% of FVIII), and two C domains. However, the B domain is not necessary for the coagulation function of FVIII. In recent decades, researchers have developed a modified version of FVIII, called B-domain deleted FVIII (BDD-FVIII), which lacks most of the B domain and is less than 5 kb in size. This smaller version can be more effectively packaged into AAV vectors for gene therapy (Batty and Lillicrap, 2019).

2.6 Genome editing

Genome editing refers to a collection of techniques employed to alter the DNA sequence of a cell. This approach represents a distinct subtype of gene therapy, by which it is possible to generate a novel category of therapeutics capable of treating genetic and non-genetic ailments. The limitations of gene editing in mammalian cells can be overcome by inducing double-strand breaks (DSB) in the genome, which recruit the endogenous repair machinery and can be repaired by either non-homologous end joining (NHEJ) pathways or homology-directed repair (HDR)

pathways. Meganucleases, which are low frequency cutters with a fixed recognition sequence, were initially used to induce DSB, but their limited flexibility in targeting specific loci led to the re-engineering of their DNA recognition site. Zinc fingers, small protein motifs with sequence-specific DNA binding capacity, were discovered to be highly versatile in binding specific DNA sequences of various lengths. Zinc finger nucleases (ZFN) were generated by fusing zinc finger modules to the DNA cleavage domain of the restriction enzyme FokI, which could induce DSB and subsequently enhance HDR. DSB caused by ZFNs can be repaired by HDR if a repair template is provided, which allows for knock-out (KO) and knock-in (KI) mutations to be generated with genome editing. In 2010, transcription activator-like effector nucleases (TALENs) were generated by exchanging the DNA binding domain of ZFN with more flexible and easier to generate DNA binding modules derived from TALE proteins in plant pathogenic bacteria. TALENs were found to require less effort and be more flexible in design and generation than ZFNs, as each peptide in the TALEN module can bind to a single nucleotide unlike ZNF were a single peptide binds to 3 nucleotides (64 combination) (Porteus, 2019; Tröder and Zevnik, 2022).

The possibility of a protein with programmable DNA binding capabilities that does not require the time-consuming construction of a new protein domain has been an attractive goal in biotechnology. In nature, such a system was discovered and is known as the CRISPR/Cas system in bacteria. The CRISPR/Cas system is a defence mechanism against infections and is similar to the adaptive immune response known from higher organisms such as mammals. The system uses CRISPR-associated or Cas genes, which are naturally occurring endonucleases. In addition, the system uses CRISPR RNA (crRNA) transcribed from the spacer to guide the Cas protein to its target. The target is DNA, and a short sequence motif adjacent to the crRNA targeted

sequence on the target DNA, called the protospacer adjacent motif (PAM), is critical for cleavage and responsible for self vs. non-self-discrimination of the CRISPR system (Adli, 2018).

CRISPR/Cas9 system is a type of RNA-guided nuclease that has revolutionized the field of genome editing. The system was initially developed using the Cas9 protein from Streptococcus pyogenes, which has a simple PAM requirement of NGG. However, researchers have developed other Cas9 and Cas9-like proteins with different properties, including alternative PAM requirements, protein size, and substrate specificity, to expand the applicability of the system. Protein engineering has led to the development of various Cas proteins with altered PAM requirements, expanding their targeting scope to previously inaccessible regions. The large protein size of SpCas9 prevents its delivery using the adeno-associated virus system, therefore a smaller Cas9 variants like SaCas9 were employed. The discovery of Cas12a introduced an entirely new class of CRISPR proteins that can target TT-rich PAM targets. The CRISPR/Cas13 system, an RNA editing system, can be used to modify the function of a gene by degrading the mRNA, reducing a gene's function by knockdown instead of complete removal. Cas9 variants with modified nucleases, such as dCas9 which can bind to but not cleave nucleic acid by inactivating the nuclease activity of Cas9 through two-point mutations in the HNH and RuvC domains, and Cas9 nickases which cut a single strand by mutating only one of the two nucleases were created (Prasad *et al*., 2021).

2.7 Evolution of CRISPR Cas9 based tools

The CRISPR Cas9 system is a powerful genetic tool that allows researchers to modify specific DNA sequences in a wide range of organisms. This technology has evolved rapidly since its discovery in the 2012, with researchers developing

various Cas9 protein variations, such as Cas12a and Cas13, for different functions beyond genome editing. These tools can now be used for gene activation, repression, and epigenome editing, broadening the scope of applications for CRISPR Cas9 in research and biotechnology. Improvements have also been made to enhance the precision and specificity of CRISPR Cas9 tools to reduce the risk of off-target effects. Overall, the evolution of CRISPR Cas9 based tools has revolutionized the field of genome editing and opened up new avenues for research in genetics, medicine, and biotechnology (Li *et al*., 2023).

2.7.1 based on double strand break

The CRISPR/Cas system causes breaks in both strands of DNA, which can be repaired through either Non Homologous End Joining (NHEJ) or homology directed repair (HDR). NHEJ repairs the broken strands by introducing small, random insertions or deletions, resulting in mutations that can permanently knock out the target gene, although efficiency may vary depending on the site. Targeting two sites simultaneously with appropriate gRNAs can also lead to deletion or inversion of the sequence in between the two targets. HDR, on the other hand, relies on a donor carrying a gene of interest or a desired mutation to be inserted at the break site with homology arms on both sides. This method is less efficient and more complex than NHEJ but offers seamless editing. The homologous DNA strand can be delivered as single-stranded oligonucleotide or double-stranded DNA through transfection or transduction. HDR can be used to insert the desired gene or epitope tag at the targeted site. Various methods were optimized to increase HDR efficiency, such as using small molecules to stall cell division, inhibiting NHEJ enzymes, creating staggered cuts, and fusing HDR repair proteins to Cas9 (Prasad *et al*., 2021).

2.7.2 based on Cas9 nicking

The development of base editors was made possible by the fusion of other proteins with Cas9, which allows for the conversion of single nucleotides without causing double strand breaks. The first base editor, called cytosine base editor (CBE), was created by fusing APOBEC1, an enzyme in the cytosine deaminase family, to Cas9 nickase. The resulting editor can convert C.G to T.A by positioning the cytosine deaminase in the correct orientation on the target DNA strand, and deaminating cytosine to uracil. The uracil glycosylase inhibitor (UGI) fused to Cas9 prevents the conversion of U-G back to C-G by base excision repair. Another base editor, called adenosine base editor (ABE), was created by evolving the bacterial enzyme TadA to convert adenosine to inosine, which is subsequently converted to guanosine. ABEs showed better efficiency and less off-target effects compared to CBEs. Development of base editors expanded the application of the CRISPR/Cas system, allowing for mutations to be created without causing any double strand break or significant amount of indels. Over time, several groups worked on improving the efficiency of both base editors. Recently, scientists also developed CGBE which is able to induce C to G transversion, thus expanding the number of genetic diseases that can be corrected using base editing. EvolvR is a genome editing tool that can continuously introduce mutations in the region targeted by the Crispr/Cas system, while Prime editing allows for single base conversion, insertions or deletions at the target site with reduced PAM constraints (Prasad *et al*., 2021).

2.7.3 based on dCas9

To achieve precise control over gene expression, a mutant form of the Cas9 enzyme called Dead Cas9 (dCas9) was employed. Unlike the wild-type Cas9, dCas9

lacks the ability to induce double-strand breaks in DNA. Instead, it functions as a gene knockdown tool by tightly binding to the target DNA sequence, thereby obstructing the activity of other DNA-binding proteins and impeding transcription. When combined with a potent repressor complex like KRAB, dCas9 generates a stronger and more targeted gene repression effect. Conversely, the recruitment of robust transcriptional activators using the dCas9 targeting platform leads to robust induction of gene expression. The choice of the optimal method is likely to depend on the specific cell type and experimental context. In recent comparative studies encompassing multiple species, the VPR, SAM, and SunTag systems have consistently outperformed the traditional VP64 system. These advanced CRISPR-based systems have provided researchers with enhanced capabilities to achieve precise spatial and temporal control over the dynamics of gene expression from a target locus. Through the utilization of various strategies facilitated by the CRISPR methodology, significant progress has been made in accomplishing the goal of attaining precise regulation of gene expression (Prasad *et al*., 2021).

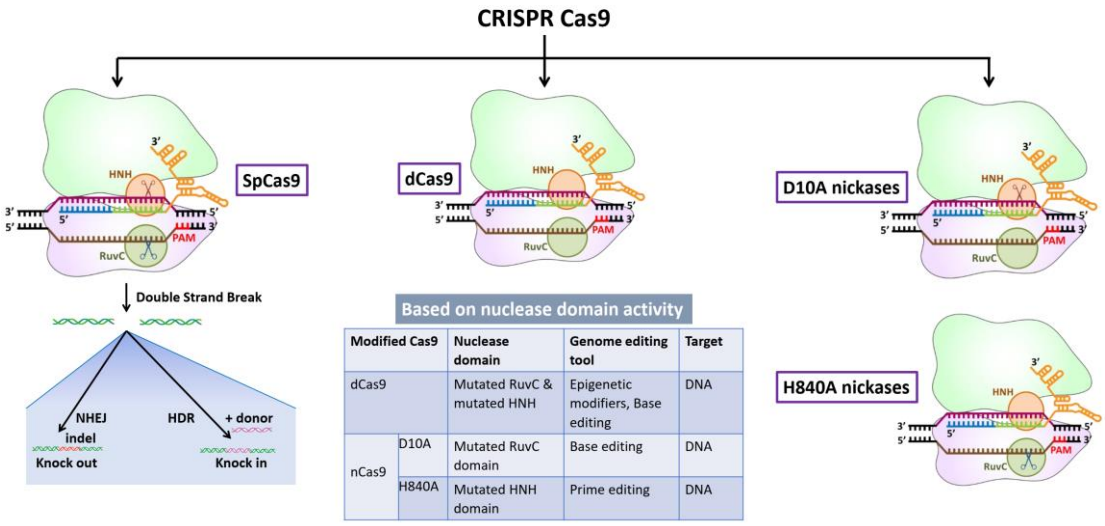


Figure 2.4: Harnessing the Nuclease Domain in CRISPR-Cas9 for Advanced Genome Editing Tools

2.8 Base editing

The use of nucleases such as Cas9 in genome editing results in undesired genetic changes, such as indels, translocations, and rearrangements. Base editing is an alternative method that can directly generate precise point mutations in genomic DNA or RNA without causing DSBs. By avoiding DSBs, base editors reduce the formation of DSB-associated byproducts. DNA base editors operate on single-stranded DNA and do not modify double-stranded DNA. They use a deaminase enzyme to modify DNA bases within a single-stranded DNA bubble created by the displacement of a small segment of single-stranded DNA by the guide RNA. The catalytically disabled nuclease used in base editing also generates a nick in the non-edited DNA strand to facilitate repair using the edited strand as a template. There are two classes of DNA base editors: cytosine base editors (CBEs) and adenine base editors (ABEs). CBEs convert a C•G base pair into a T•A base pair, while ABEs convert an A•T base pair to a G•C base pair, enabling all four possible transition mutations. In addition, RNA base editors that target adenosine conversion to inosine have been developed using both antisense and Cas13-guided RNA-targeting methods (Liu, 2018).

2.8.1 Cytosine base editor (CBE)

Cytosine base editor was the first DNA base editors which can change a C•G base pair to a T•A base pair by removing the exocyclic amine from the target cytosine and producing uracil. To focus deamination activity on a small target window within the mammalian genome, APOBEC1 cytidine deaminase was fused to dead Cas9 from Streptococcus pyogenes (dCas9) to create base editor 1 (BE1). However, BE1 is not effective in human cells because of effective cellular repair of the U•G intermediate in DNA. To address this, uracil DNA glycosylase inhibitor (UGI) was fused to BE1, creating BE2, which inhibits UNG and enables base editing in mammalian cells.

Third-generation base editors (BE3) were designed to specifically nick the non-edited DNA strand to direct cellular replacement of the G present in the non-targeted strand of DNA with A. Target-AID is a similar system for cytosine base editing in yeast and mammalian cells that uses the cytidine deaminase CDA1 in a Cas9 nickase–CDA1–UGI base editor construct. However, at specific genomic locus, the use of cytosine base editor results in reduced editing efficiency and lower product purity of the desired edits (Komor, Kim, Packer, Zuris and David, 2016).

Several strategies have been implemented to improve the editing efficiency of cytosine base editors (CBEs). These techniques involve optimizing codons to the human system, utilizing phage-assisted evolution, and employing directed evolution methods. The codon optimization of base editors entails altering their sequence to better align with the availability of amino acids in the human system, which can improve its performance and reduce immunogenicity. Additionally, directed evolution has been employed to optimize the CBE, which involves the generation of a diverse library of CBE variants with random mutations, followed by selection of those variants with improved properties. Phage-assisted evolution has been used to further refine the CBE variants, where bacteriophages are used to infect and replicate the variants, selecting for those with the desired traits. Recent studies have also focused on improving the efficiency CBE by evolving the Tad domain used in the ABE system, using similar techniques of directed evolution and phage-assisted evolution. These approaches can be effective in increasing editing efficiency and reducing off-target effects, making them valuable tools for genome editing applications (Kim *et al*., 2017; Lam *et al*., 2023; Neugebauer *et al*., 2022).

2.8.2 Adenine Base Editor (ABE)

Adenine also contains an exocyclic amine that can be deaminated to alter its base pairing preferences, and deamination of adenosine yields inosine. Inosine in the context of a polymerase active site exhibits the base-pairing preference of guanosine. The major hurdle to developing an adenine base editor was the lack of any known adenosine deaminase enzymes capable of acting on ssDNA. However, adenosine deaminase enzyme targeting tRNA of Escherichia coli was evolved. Through directed evolution E. coli cells were equipped with TadA mutants and defective antibiotic resistance genes. To grow in the presence of antibiotic, a mutant TadA–dCas9 fusion (TadA*–dCas9) must convert a deoxyadenosine to a deoxyinosine in the defective antibiotic resistance gene. Bacteria encoding TadA–dCas9 fusions capable of repairing the mutated resistance gene were isolated and then tested in a mammalian cell context. Simple TadA*–Cas9 nickase fusions resulted in only modest editing rates in mammalian cells, and this challenge was addressed by engineering heterodimeric proteins that incorporate a wild-type non-catalytic TadA monomer, an evolved TadA* monomer, and a Cas9 nickase (TadA–TadA*–Cas9 nickase) in a single polypeptide chain. As with CBEs, ABEs catalyze deamination within a small window of exposed ssDNA generated by Cas9:guide RNA binding to the target locus. Thus ABE7.10 was evolved, containing 14 amino acid substitutions in the catalytic TadA* domain (Nicole M. Gaudelli *et al*., 2017).

The efficiency of ABE7.10 for genome editing purposes was improved using two techniques: phage assisted evolution and directed evolution. Phage assisted evolution involved selecting for desirable mutations in a population of cells using bacteriophages, resulting in the creation of ABE8e. Directed evolution introduced random mutations into the protein sequence of ABE7.10, followed by selection for

desirable traits, leading to the creation of ABE8.16. Both ABE8e and ABE8.16 demonstrated higher efficiency and lower off-target effects compared to ABE7.10, indicating improved accuracy and precision in genome editing. These advancements have significant implications for genetic research and therapy by providing more effective tools for manipulating DNA sequences with precision and accuracy (Richter *et al*., 2020).

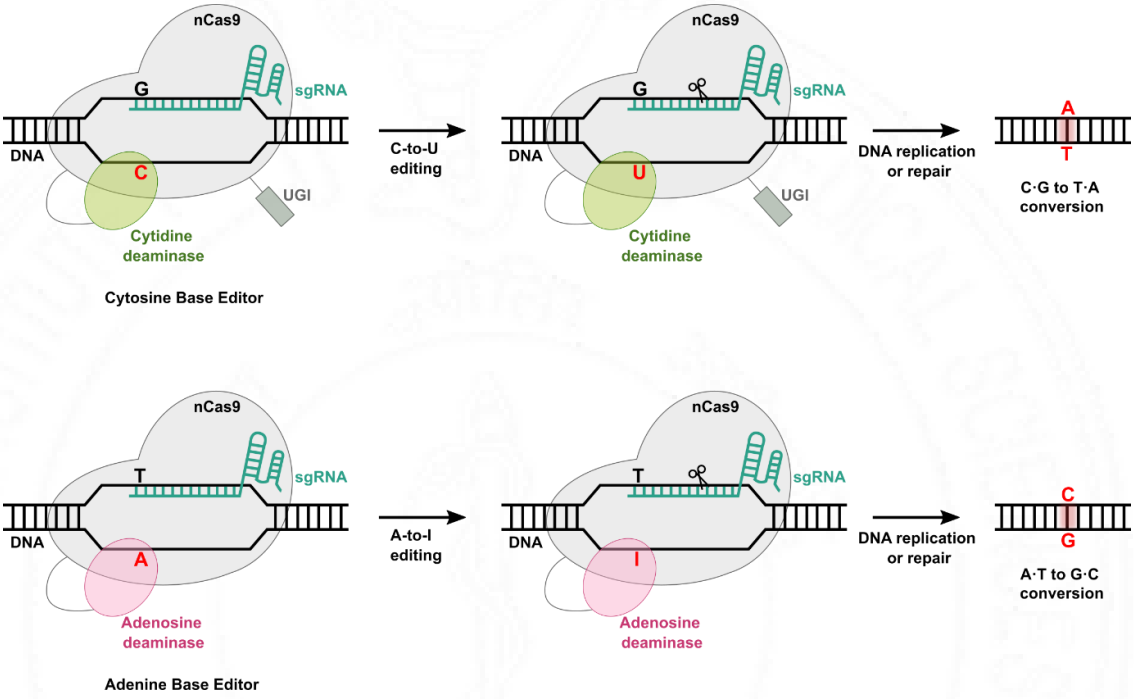


Figure 2.5: Molecular Mechanisms of Nucleotide Substitution by DNA Base Editors: Cytidine Base Editors (CBEs) and Adenine Base Editors (ABEs)
(Source - https://biotech.ucdavis.edu/news/dna-base-editors-genome-editing)

2.8.3 Others

RNA base editing is a powerful tool for the life sciences and medicine, allowing the editing of individual bases in RNA. These technologies rely on using adenosine deaminases from the ADAR family, which convert adenosine to inosine, to edit RNA. The RNA editors are guided to their target RNA by antisense oligonucleotides that recognize and localize the ADAR deaminase domain to certain regions of double-stranded RNA. Various strategies have been developed to establish

a physical linkage between the deaminase and the antisense RNA. Innovations have improved the efficiency and specificity of RNA-guided deamination systems, but the methods are still limited by off-target editing and context-dependent editing of adenine bases (Monian *et al*., 2022).

To perform transversion editing using base editors, previously observed unintended C>G edits by ABE max were exploited. Introducing two uracil glycosylase inhibitor (UGI) domains to ABEmax resulted in a reduction of C-to-G edits and indels. To improve C>G editing, BEmax CBE was chosen and modified by removing the UGI domains. The resulting BE4maxΔUGI, based on the rat APOBEC1 cytidine deaminase and lacking the two UGIs, showed an increase in C-to-G (and to a lesser degree C-to-A) edits compared to wild-type BE4max in HEK293T cells. To further enhance the editing efficiency, the human UNG (hUNG) enzyme was added, which promotes more glycosylation (Koblan, Arbab, *et al*., 2021).

2.9 Homologous directed repair (HDR)

Homologous directed repair is a high-fidelity repair pathway, repairs double-strand DNA breaks (DSBs) during S-or G2-phase of cell cycle, using homologous DNA sequences. The process of HDR involves multiple steps that are tightly regulated and coordinated by a complex network of proteins. Initially, the DSB is recognized by the MRE11-RAD50-NBS1 (MRN) complex, which recruits ATM kinase to the site of the break. ATM kinase phosphorylates H2AX, leading to the formation of γH2AX foci, which marks the damaged site for repair. This signaling cascade activates downstream effectors that mediate the repair process. In the next

step in HDR involves the resection of the broken DNA ends by nucleases, such as Exo1 and DNA2, creating a single-stranded DNA (ssDNA) overhang. Replication protein A (RPA) binds to the exposed ssDNA and prevents it from forming secondary structures. RPA also recruits other proteins, such as RAD51 and BRCA1, to the site of the break. RAD51 forms a nucleoprotein filament on the RPA-coated ssDNA, which searches for a homologous sequence in the sister chromatid or homologous chromosome. The homologous sequence serves as a template to guide the synthesis of the missing DNA strand, leading to the formation of a Holliday junction. The Holliday junction is then resolved by either a crossover or a non-crossover mechanism, depending on the stage of the cell cycle and the regulatory factors involved. This process restores the original DNA sequence, with minimal or no loss of genetic information (Krejci *et al*., 2012; Li and Heyer, 2008; Sung and Klein, 2006).

CRISPR-based homology-directed repair (HDR) is a powerful tool for precise genome editing that relies on the endogenous HDR machinery to introduce specific modifications at targeted sites in the genome. The CRISPR system utilizes a guide RNA (gRNA) that directs the Cas9 endonuclease to a specific DNA sequence, where it creates a DSB. This DSB can be repaired by either non-homologous end-joining (NHEJ), which frequently results in insertions or deletions (indels), or HDR, which can introduce precise changes by using a homologous DNA template. To facilitate HDR, a donor DNA template with homology arms flanking the desired modification can be introduced into the cell along with the CRISPR components. The homology arms are complementary to the sequences adjacent to the DSB and serve as a template for DNA repair. The donor DNA can carry a variety of modifications,

such as single nucleotide changes, deletions, insertions, or even large genomic rearrangements. The frequency of HDR-mediated repair can be increased by optimizing the length and orientation of the homology arms, as well as the delivery method and timing of the CRISPR components (Porteus, 2019; Sharma *et al*., 2016; Wang *et al*., 2015).

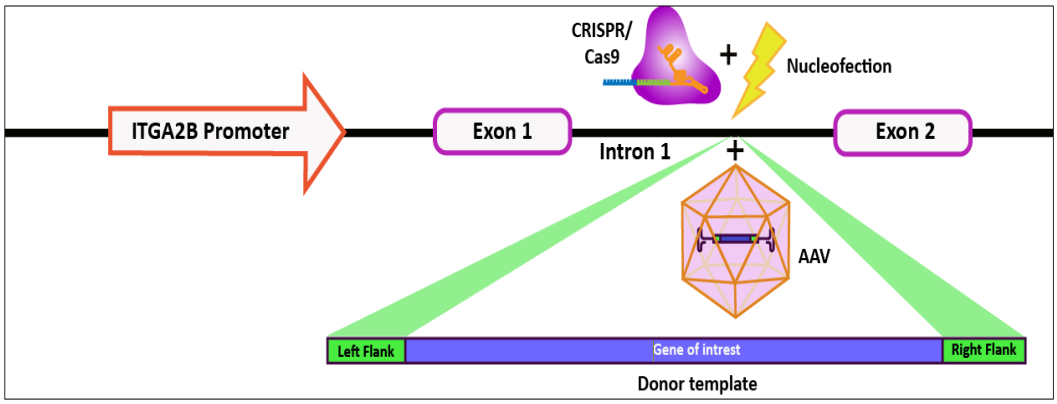


Figure 2.6: Mechanisms of Homology-Directed Repair Utilized by CRISPR/Cas9 and AAV based donor template.

2.9.1 Types of donor template

Different types of templates, including plasmid, PCR, ssODNA, and AAV-based templates, can be used for introducing the desired modification in the genome. Plasmid-based templates are circular DNA molecules that can be easily manipulated in the laboratory and used to introduce large DNA fragments. These templates contain the desired modification flanked by homology arms that facilitate recombination with the target site in the genome. PCR-based templates are smaller in size than plasmid-based templates and can be easily synthesized in the laboratory. The PCR product contains the desired modification flanked by homology arms that facilitate recombination with the target site in the genome. Single-stranded oligonucleotides (ssODNA) are short DNA fragments that are highly efficient in

CRISPR HDR due to their small size and easy synthesis in the laboratory. The ssODNA templates contain the desired modification and homology arms that can facilitate recombination with the target site in the genome. AAV-based templates utilize adeno-associated virus (AAV) vectors to deliver the desired modification and homology arms to the target site in the genome. AAV-based templates have been shown to be highly efficient in CRISPR HDR. The choice of template depends on the size of the desired modification and the specific application (Bak *et al*., 2017; Bijlani *et al*., 2022; Dewitt *et al*., 2017; Richardson *et al*., 2016).

Although different types of templates, such as plasmid, PCR, ssODNA, and AAV-based templates, have been used for CRISPR HDR, each of these methods has limitations and drawbacks. For plasmid-based templates, their large size results in inefficient delivery. For PCR-based templates, shorter homology arms and high error rates reduce their efficiency in HDR. Additionally, ssODNA-based templates are susceptible to degradation and are not suitable for introducing large modification. Moreover, the delivery of naked DNA into cells through transfection or electroporation can trigger an immune response, resulting in cellular toxicity. AAV-based templates have limited cargo capacity and may also elicit immune responses in certain applications. However, several modifications have been proposed to overcome these limitations. For example, the use of chemical modifications such as 2'-O-methyl and phosphorothioate linkages can improve the stability and efficiency of ssODNA-based templates. In addition, using smaller homology arms and self-complementary AAV vectors can increase the cargo capacity and reduce immune responses in AAV-based templates (Gaj *et al*., 2017; Shakirova *et al*., 2023).

2.9.2 homology arm modifications

The DNA regions flanking the desired insertion site that are complementary to the target sequence are referred to as "homology arms." Homology arms facilitate the incorporation of exogenous DNA sequences and guide the repair machinery to the desired location in homology-directed repair (HDR) gene editing.

In mammalian cells, homology-directed repair (HDR) is not very efficient, whereas non-homologous end joining (NHEJ)-mediated repair is the primary pathway utilized for repair. To overcome the limitations of genome editing using homology-directed repair (HDR), a homology-independent targeted integration (HITI) technique was developed. HITI uses CRISPR/Cas9 system to insert the donor template efficiently into the genomic double-strand break (DSB) site in both dividing and non-dividing cells *In Vitro* and in vivo. Unlike HDR, HITI relies on the non-homologous end joining (NHEJ) repair pathway. In HITI, the donor DNA lacks a homology arm and is designed to include a Cas9 cleavage site flanking the donor sequence. Cas9-mediated DSBs are created simultaneously in both genomic target sequences and the exogenous donor DNA, leading to the formation of blunt ends. The linearized donor DNA can then undergo repair by the NHEJ pathway, allowing for its integration into the genomic DSB site. Once incorporated, the donor DNA inserted in the desired orientation disrupts the Cas9 target sequence and prevents further Cas9 cutting (Bloomer *et al*., 2021; Suzuki *et al*., 2016).

DNA-repair complexes have greater flexibility than previously believed and are not confined to only the NHEJ or HDR pathways. These complexes have been found to operate even in post-mitotic cells. To address the limitations of existing gene-editing methods, a novel approach called intercellular linearized Single

homology Arm donor mediated intron-Targeting Integration (SATI) was developed. SATI requires a donor DNA molecule that includes a single stretch of homologous sequence at the DSB induction site. This method enables the targeting of a wide range of mutations and cell types, as it allows for both homology-directed repair (HDR) and homology-independent targeted integration (HITI) via the use of a single homology arm (Suzuki *et al*., 2019).

2.9.3 Small molecules to enhance HDR

Strategies to enhance HDR-mediated gene editing include inhibition of the competitive NHEJ pathway, activation of HDR-promoting factors, altering the chromatin state, and altering the cell cycle of HSPCs. Small molecules such as SCR-7, NU7441, NU7026, STL127685, IC86621, M3814, and KU-0060648 are tested for NHEJ inhibition. L755507, resveratrol, brefeldin A, RS-1, MLN4924, and NSC 15520 are tested for enhancing HDR. HDAC inhibitors such as Entinostat, panobinostat, TSA, VPA, and romidepsin are employed to modulate chromatin state. Restricting gene editing to S/G2 phase by Cas9-geminin fusion, XL-413, an inhibitor of CDC7, nocodazole, aphidicolin, and Rapamycin and CHIR99021 were tested for altering the HSPC cell cycle. However, the efficiency of small molecules varies based on the cell types and gene locus (Lin *et al*., 2014).

2.10 Delivery method

The efficacy and safety of CRISPR gene therapy heavily rely on the chosen delivery method. Despite concerns regarding the potential risks associated with classical virus-based gene therapy, adeno-associated virus (AAV) vectors continue to be widely used in CRISPR gene therapy due to their high delivery efficiency.

Alternatively, the CRISPR system can be delivered using various techniques, including electroporation, lipid nanoparticles, microinjection of plasmid DNA, mRNA, or protein, as well as incorporation into AAV vectors for targeted cell delivery. Electroporation has the drawback of potential cell toxicity and long-term permeabilization, while microinjection is technically challenging and more suitable for ex vivo applications. AAV-based delivery may lead to prolonged expression of CRISPR components, increasing the risk of off-target effects. Conversely, the administration of Cas9 protein and guide RNA (gRNA) as ribonucleoprotein (RNP) complexes has shown reduced off-target effects and sustained editing efficiency due to their temporary expression and rapid clearance from cells (Yip, 2020).

Currently, the Cas9 RNP complex and AAV vectors are the two most commonly utilized approaches for in vivo distribution of CRISPR/Cas9 gene editing tools. Depending on the chosen delivery strategy, gene editing can be performed in vivo or ex vivo. Ex vivo editing involves extracting target cells from the patient, modifying them outside the body, and reintroducing them afterward. This approach offers improved safety, technical feasibility, and stricter quality control over modified cells. However, challenges include maintaining in vivo cell functionality post-genetic modification, intensive in vitro cell culture, and obtaining a sufficient number of cells for effective re-engraftment. In contrast, in vivo editing allows for the treatment of a wider range of genetic disorders but requires systemic or localized delivery of CRISPR therapeutics. In vivo delivery poses various challenges, including degradation by circulating proteases or nucleases, opsonization by opsonins, clearance by the mononuclear phagocyte system, and the need to bypass the vascular endothelium. Additionally, when the editing apparatus is localized near the injection site, the distribution of modified cell population within the tissue may be uneven.

While direct delivery is preferred over viral vectors for in vivo therapy, concerns have prompted researchers to explore alternative non-viral delivery approaches for in vivo gene editing, such as nanoparticles and liposomes (Yip, 2020).

2.11 HSPCs preferred cells for editing hereditary hematological disorders

Hematopoietic stem and progenitor cells (HSPCs) possess a high therapeutic potential due to their ability to both self-renew and differentiate. A small number of genetically modified HSPCs could provide a lifelong, corrective reconstitution of the hematopoietic system in patients with various hematologic disorders. Gene addition therapies targeting HSPCs using integrating retroviral vectors have been shown to have clinical benefits in immunodeficiencies, storage disorders, and hemoglobinopathies. However, incomplete knowledge of target cells, genomic control, and gene transfer technologies have led to unexpected hurdles in developing clinically relevant gene therapy protocols. To address these challenges, targeted gene-editing technologies using engineered nucleases such as ZFN, TALEN, and/or CRISPR/Cas9 RGEN are being developed. These technologies offer a promising approach to correct disease-causing mutations with specific applications in autosomal dominant or dominant-negative genetic disorders (Yu *et al*., 2016).

Hematopoietic stem/progenitor cell (HSPC) transplantation has the potential to treat a wide range of congenital and hereditary diseases by replacing some or all of the patient's bone marrow with donor cells. However, autologous transplantation with gene therapy is proposed as an alternative treatment for patients who cannot find an HLA-matched donor. Despite the potential benefits, there are several challenges to overcome before HSPC gene editing can be widely used in clinical applications. One major challenge is the high sensitivity of primitive long-term repopulating

HSPCs to DNA damage caused by genome-editing tools. The particularly poor HDR rate observed in primitive HSPCs is likely due to quiescence or slow cycling of these cells. Therefore, shortening the length of exposure to nuclease activity may help eliminate this issue, but the benefits must be balanced against the need for efficient editing. Moreover, the levels of genome editing and cell viability are inversely correlated, and cell viability is closely correlated with engraftment rate. To optimize the delivery method used to transport gene-editing tools into HSPCs, it is necessary to enrich for edited HSPCs and *In Vitro* expand them to ensure a safe dose of engrafting cells before transplantation (Yu *et al*., 2016).

2.12 Safety

Genome editing is a rapidly advancing field with numerous potential applications in various fields of medicine. One of the primary concerns with genome editing is the possibility of off-target effects, where unintended genomic modifications occur, leading to unpredictable outcomes. The potential for engraftment, multilineage potential, viability, and differentiation are also important factors to consider when assessing the safety of genome editing. Research in these areas is ongoing, and the development of new techniques and technologies is continually improving the safety and efficacy of genome editing. While there are still concerns regarding the safety of genome editing, with continued research and refinement of techniques, genome editing has the potential to revolutionize the field of medicine (Tao *et al*., 2023).

Nucleases and repair templates, analysis of off-target editing, and careful utilization of viral vectors can all be used to improve safety. Gene editing tools including zinc finger nucleases, homing endonucleases, TAL effector nucleases and RNA-guided nucleases allow for the targeted modification of cellular genomes.

Repair of the double-strand breaks generated by these tools occurs through non-homologous end joining or homology directed repair mechanisms. When genome editing is used therapeutically, controlled genetic damage is generated, and the native repair mechanisms of the cell repair the damage. The most fundamental safety concern with gene editing is the potential for unintended double-strand breaks. When designing an endonuclease, it is important to anticipate potential off-target cleavage sites and select sequences that minimize this potential.

The primary strategies that have been successfully used to reduce off-targeting in CRISPR-Cas gene editing can be grouped into three main categories: (I) using bioinformatics tools to design more accurate gRNA and predict off-targeting; (II) utilizing Cas9 nickases; and (III) adding anti-CRISPR proteins. Bioinformatics tools play a critical role in the CRISPR Cas system by allowing for the analysis, prediction, and determination of the system. These tools help to design more efficient gRNAs and detect the accurate editing site within the genome, which in turn reduces off-targeting. Cas9 nickase is another practical approach that helps to reduce off-targeting by mutating one nuclease domain in just one strand of the DNA. This creates a nick that is quickly repaired by the cells. Finally, anti-CRISPR proteins are used to inactivate the Cas9 protein after targeting its site, thereby reducing off-targeting.

Engraftment potential is an important consideration in genome editing, as it determines the ability of edited cells to survive and proliferate in the body after transplantation. Several studies have demonstrated the ability of genome-edited cells to engraft in vivo. However, the efficiency and long-term stability of engraftment can vary depending on the cell type, the editing technique used, and the host environment. For example, some cell types may be more amenable to engraftment than others.

Additionally, certain editing techniques may be more effective at promoting engraftment, such as those that minimize DNA damage or that promote the repair of DNA breaks. Furthermore, the host environment can also influence engraftment, as factors such as the immune system and the availability of growth factors and nutrients can affect the survival and proliferation of transplanted cells.

Multilineage potential is also an important consideration in genome editing, particularly for clinical applications that require the generation of specific cell types. It is crucial that edited cells have the capacity to differentiate into the intended cell type without causing any changes to their ability to differentiate after editing. Studies have shown that genome-edited cells can maintain multilineage potential. However, the efficiency and stability of differentiation can vary depending on many factors, including the editing technique used, the cell type, and the host environment. Therefore, careful consideration of these factors is necessary to optimize the generation of desired cell types.

Viability is another important consideration in genome editing, as it determines the ability of edited cells to survive and function normally after genome editing. Genome editing can reduce cell viability, particularly if the editing efficiency is low or if the editing leads to DNA damage. However, new techniques, such as CRISPR/Cas9 ribonucleoprotein delivery, have been developed to improve viability and minimize DNA damage. These techniques can increase the efficiency of genome editing while minimizing the risk of deleterious effects on cell viability.

MATERIALS AND METHODS

SECTION 3.1: IDENTIFICATION OF NOVEL HPFH-LIKE MUTATIONS BY CRISPR BASE EDITING THAT ELEVATE THE EXPRESSION OF FETAL HEMOGLOBIN.

3.1.1 Designing and cloning of the gRNA

The gRNAs were created for targeting the *HBG1* and *HBG2* promoter region. SnapGene and Benchling software were used to design the gRNAs. The gRNAs for CBE were selected from 43 hits using the "gRNAs for base editing" design type in the Benchling tool, and 32 non-overlapping guides were chosen. The gRNAs for ABE were manually designed using SnapGene software. The forward oligonucleotide contained the gRNA sequence without PAM and "CACCG" overhang at the 5' end, while the reverse oligonucleotide contained the reverse complement of gRNA without PAM, "AAAC" overhang at the 5' end, and a "C" added at the 3' end. The complementary oligonucleotides listed in Table 3.1.1 were annealed and cloned into BsmBI digested pLKO5.sgRNA.EFS.GFP/RFP vector (Addgene #57822/#57823(Heckl *et al*., 2014; Ran *et al*., 2013; Shalem *et al*., 2014) The annealed products were diluted 1:200 fold, and 6 µl was taken along with 50 ng of vector backbone for the ligation reaction, as per the manufacturer's instruction from NEB. The ligated product was transformed into DH10B competent cells and plated in LB-agar containing 100 µg/ml of Ampicillin for selection (Sambrook and Russell, 2001). Three colonies were picked and inoculated in LB broth for colony PCR, which was carried out using GoTaq® Hot Start Polymerase premix and sequencing primers. After confirming the expected amplification in 1% agarose gel, the second round of PCR was carried out using the pre-cleaned product from the first round of PCR using the BigDye™ Terminator v3.1 Cycle Sequencing Kit, as per the manufacturer's protocol, and sent for Sanger sequencing. The primers used for confirming the clones are provided in Table 3.1.2.

Table 3.1.1: The gRNAs used in this study to screen the *HBG* promoter region and their respective primer for sequencing.

Sl. No.	Names	gRNA Info 5'-3'	Editing window for CBE	Editing window for ABE	Primers used for Sanger sequencing	Primers used for NGS	Primers used for HBG1/HBG2 separation
1	gRNA 1	ggctagggatgaagaataaa	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
2	gRNA 2	cttgaccaatagccttgaca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
3	gRNA 3	atAtttgcattgagatagtg	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
4	gRNA 4	gtggggaaggggcccccaag	nil	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
5	gRNA 5	tggtcaagtttgccttgtca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
6	gRNA 6	gtttgccttgtcaaggctat	yes	nil	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
7	gRNA 7	cttgtcaaggctattggtca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
8	gRNA 8	caaggctattggtcaaggca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
9	gRNA 9	gctattggtcaaggcaaggc	nil	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
10	gRNA 10	aggcaaggctggccaaccca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
11	gRNA 11	ggcaaggctggccaacccat	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
12	gRNA 12	aaggctggccaacccatggg	yes	nil	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
13	gRNA 13	cccatgggtggagtttagcc	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
14	gRNA 14	ccatgggtggagtttagcca	nil	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
15	gRNA 15	gctaaactccacccatgggt	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R

16	gRNA 16	ccctggctaaactccaccca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
17	gRNA 17	tatctgtctgaaacggtccc	yes	nil	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
18	gRNA 18	tatttgcattgagatagtgt	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
19	gRNA 19	atgcaaatatctgtctgaaa	yes	Yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
20	gRNA 20	ggaatgactgaatcggaaca	yes	yes	HBF 2 F/HBF1R	NGS 3 F/R	NGS4 F/NGS 2 R
21	gRNA 21	actgaatcggaacaaggcaa	yes	yes	HBF 2 F/HBF1R	NGS 3 F/R	NGS4 F/NGS 2 R
22	gRNA 22	aaaaactggaatgactgaat	yes	yes	HBF 2 F/HBF1R	NGS 3 F/R	Not applicable
23	gRNA 24	gcattgagatagtgtgggga	nil	yes	HBF 2 F/HBF1R	NGS 2 F/R	Not applicable
24	gRNA 25	attgagatagtgtggggaag	nil	yes	HBF 2 F/HBF1R	NGS 2 F/R	Not applicable
25	gRNA 29	agaataaattagagaaaaac	nil	yes	HBF 2 F/HBF1R	NGS 3 F/R	Not applicable
26	gRNA 30	ggagaaggaaactagctaaa	nil	yes	HBF 2 F/HBF1R	NGS 3 F/R	NGS4 F/NGS 2 R
27	gRNA 32	cagttccacacactcgcttc	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
28	gRNA 33	cttcatccctagccagccgc	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
29	gRNA 34	cctagccagccgccggcccc	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
30	gRNA 35	ccgccggcccctggcctcac	yes	nil	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
31	gRNA 36	actggatactctaagactat	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
32	gRNA 37	ccaggggccggcggctggct	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
33	gRNA 38	tgaggccaggggccggcggc	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable

34	gRNA 39	ttagagtatccagtgaggcc	nil	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
35	gRNA 40	tagtcttagagtatccagtg	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
36	gRNA 41	tagagtatccagtgaggcca	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
37	gRNA 42	agagtatccagtgaggccag	yes	yes	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
38	gRNA 43	ccagtgaggccaggggccgg	nil	yes	HBF 1 F/HBF1R	NGS 2 F/R	Not applicable
39	gRNA 44	caggggccggcggctggcta	yes	nil	HBF 1 F/HBF1R	NGS 2 F/R	NGS4 F/NGS 2 R
40	gRNA 45	aagcagcagtatcctcttgg	yes	yes	HBF 2 F/HBF1R	NGS 3 F/R	Not applicable
41	gRNA 46	attaagcagcagtatcctct	yes	yes	HBF 2 F/HBF1R	NGS 3 F/R	Not applicable
42	Control	Plasmid without gRNA	-	-	HBF 1 F, 2 F/HBF1R	NGS 2 and NGS 3	NGS4 F/NGS 2 R
43	AAVS1	ggggccactagggacaggat	yes	yes	Table 8	Not applicable	Not applicable
44	CCR5	aaaatataatctttaagata	no	yes	Table 8	Not applicable	Not applicable

Table 3.1.2: Primers to confirm the gRNA clones.

Sl. No	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	Sequencing F	gagggcctatttcccatgat	539bp	This study
2	Sequencing R	tggatctctgctgtccctgt		This study

3.1.2 Plasmid Constructs:

In this study, two base editing plasmids were predominantly used: pLenti-FNLS-P2A-Puro (Addgene#110841-CBE) and pLenti-ABERA-P2A-Puro (Addgene#112675-ABE) (Zafra *et al*., 2018). For lenti virus production, 2nd generation lentiviral packaging constructs pMD2.G and psPAX2 (Addgene #12259, 11260) were used. ABE8e a hyper variant of ABE (pLenti-ABE8e-puro vector) was created by amplifying ABE8e from the TadA-8e V106W plasmid using the primers listed in Table 3. 1.3 (Richter *et al*., 2020). The PCR product was then cloned into the pLenti-ABERA-P2A-Puro backbone using HIFI assembly (NEB) after digestion with BamH1 and Nhe1. For the construction of Lenti-based Cas9 system, lentiCRISPR V2 vector (Addgene#52961) was modified (Sanjana *et al*., 2014), the gRNA sequence was removed from the lentiCRISPR V2 vector using EcoR1 and Kpn1 enzymes, and the resulting plasmid was treated with NEB Exonuclease 1 before being ligated to produce the lentiCRISPR V2.1 vector. The plasmids were isolated using the NucleoBond Xtra Midi EF kit from Macherey-Nagel, following the manufacturer's instructions.

Table 3.1.3: ABE 8e lenti plasmid cloning primers.

Sl. No	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	ABE 8e F	acacaggtgtcgtgacgcgggatccgccacc atgaaacggacagccgacgg	4872bp	This study
2	ABE 8e R	agttggtggcgccgctgccgctagcgactttc ctcttcttcttgg		This study

3.1.3 Cell culture:

The HUDEP-2 cell lines were grown in StemSpan™ SFEM II medium which contained SCF, EPO, Pen-Strep, dexamethasone, doxycycline, and L-Glutamine. The cells were cultured at 37°C with 5% CO2 and were confirmed negative for mycoplasma. The K562 cell line was grown in RPMI supplemented with penicillin-streptomycin-glutamine and fetal bovine serum. The COS-7 and HEK 293T cells were cultured in DMEM supplemented with FBS and Pen-Strep. The confirmation for negative mycoplasma was carried out using the Universal Mycoplasma detection kit-ATCC.

Peripheral blood mononuclear cells (PBMNC) were obtained from a healthy donor and purified by density gradient centrifugation and RBC lysis. CD34+ cells were isolated from the purified PBMNC and expanded in HSC expansion media. After 24 hours of expansion, the isolated cells were analyzed for primitive cell surface markers CD34+ CD133+ and CD90+ (Genovese *et al*., 2014). The concentration of the media compositions used for HUDEP-2 and CD34+ cells are described in tables 3.1.4 and 3.1.5 below.

Table 3.1.4: HUDEP-2 expansion media:

Components	From	Stock Conc.	Working Conc.	For 10 ml	For 40 ml
StemSpan™ SFEM II	STEMCELL Technologies			10ml	40ml
SCF	Peprotech	(100ng/ul)	(100ng/ml)	5ul	20ul
EPO	Zyrop	(2U/ul)	(3U/ul)	15ul	60ul
Dexamethasone	Sigma	(100uM)	(1uM)	100ul	400ul
Doxycyclin	Sigma	(1mg/ml)	(1ug/ml)	10ul	40ul
Glutamine	Hyclone	(100X)	(1X)	100ul	400ul
Pen/Strep	Gibco	(100X)	(1X)	100ul	400ul

Table3.1.5: CD34+ HSC expansion media:

Components	From	Stock Conc.	Working Conc.	For 10 ml	For 40 ml
StemSpan™ SFEM II	STEMCELL Technologies			10ml	40ml
SCF	Peprotech	(100ng/ul)	(240ng/ml)	24ul	100ul
FLT3	Peprotech	(100ng/ul)	(240ng/ml)	24ul	100ul
IL6	Peprotech	(100ng/ul)	(40ng/ml)	4ul	16ul
TPO	Peprotech	(50ng/ul)	(75ng/ml)	15ul	60ul
Pen/Strep	Gibco	(100X)	(1X)	100ul	400ul

3.1.4 Lentivirus production:

The HEK293T cells were cultured in a 10 cm cell culture dish until they reached 80% confluency. Then, 2.5 μg of pMD2.G (envelope plasmid), and 3.5 μg of psPAX2

(packaging plasmid) were transfected into the cells along with 4 µg (for constructs with gRNA) or 5 µg (for constructs with ABE/CBE/Cas9) of lentiviral-virus production using FuGENE-HD as per manufacturers instruction. The cells were cultured for 72 hours, the viral supernatants were collected separately at 48 hours and 72 hours and concentrated using Lenti-X Concentrator. The concentrated pellets were resuspended in 200ul of 1xPBS and stored at -80°C in aliquots.

3.1.5 Lentiviral transduction:

To introduce the desired lentivirus into HUDEP-2 or K562 cells, a 100 µL aliquot of the lentivirus, 6 µg/mL Polybrene, and 1% HEPES 1M buffer were added to 0.5 million cells in one well of a 6 well plate. The cells were then spinfected at 800g for 30 minutes at room temperature using table top centrifuge, incubated with lentivirus at 37°C for 48 hours, and then incubated in fresh medium. For the generation of stable cell lines, cells transduced with the pLenti-FNLS-P2A-puro or pLenti-ABERA-P2A-puro or pLenti-ABE8e-puro or lentiCRISPR V2 1viral vector were treated with 1 µg/mL puromycin for ten days. Transduced cells with the pLKO5.sgRNA.EFS.GFP/RFP vector were analyzed by FACS for GFP/RFP expression.

3.1.6 *In Vitro* Transcription:

To prepare the template for *In Vitro* transcription (IVT), the ABE8e plasmid (Addgene#138495) was linearized using Pme1 (NEB) and purified with PCI (Phenol-Chloroform-Isoamyl alcohol). The IVT was performed using the T7 mScript™ Standard mRNA Production System (CELLSCRIPT) with complete substitution of uridine with Pseudouridine-5'-triphosphate (Jena Bioscience) (Mahalingam *et al*., 2022). The resulting mRNA was purified and stored in aliquots of 5 µg/vial at -80°C.

3.1.7 Electroporation of CD34+ cells:

CD34+ cells were grown in HSC expansion media for 48 hours, and then around 1 million of these cells were pelleted and suspended in 19 µl of MaxCyte buffer (Hyclone). ABE 8e mRNA (5 µl) and 100 pmole of the desired gRNA (1 µl) were added to the suspension. The cells were then loaded into one well of an OC25x3 Maxcyte cuvette and electroporated using the "HSC-3" program. After electroporation, the cells were transferred to a single well of a 12 well plate and allowed to recover for 20 minutes in an incubator at 5% CO2 and 37°C. To the recovered cells, 1 ml of HSC expansion media was added and then they were grown for an additional 48 hours before any further experiments were conducted.

3.1.8 Erythroid Differentiation:

We followed a previously established protocol by Trakarnsanga *et al* with some modifications for the erythroid differentiation of HUDEP2 cells (Trakarnsanga *et al*., 2017). Briefly, after expanding the cells for 8 days, approximately 1 million edited cells were seeded in a 65mm cell culture dish along with 5ml of differentiation media, which comprised IMDM glutamax, AB serum, FBS, insulin solution human, Heparin sodium salt, Holo Transferrin, EPO, SCF, IL3, Pen-Strep, and Doxycycline. The cells were then cultured in this media for 9 days, with regular media changes on Day 3 and Day 6. On Day 6, cells were cultured in erythroid differentiation medium without Doxycycline and with 500 µg/ml of holotransferrin.

For CD34+ cells, we used a three-phase liquid culture system and followed previously described methods for erythroid differentiation and enucleation analysis by Psatha *et al* (Psatha *et al*., 2018). The first composition mentioned is designated for Phase 1 and involves changing the media on day 0 and day 3. The second composition, designated for Phase 2, does not contain IL3 and hydrocortisone, and requires media

changes on day 6 and day 9. The third composition, designated for Phase 3, does not contain IL3, SCF, and hydrocortisone and requires media changes on day 12, 15, and 18. The CD235a and CD71 markers were analyzed using FACS for the erythroid differentiation pattern obtained from HUDEP-2 cells (on day 9) and CD34+ cells (on day 21). The concentration of the components used in the erythroid differentiation media for both HUDEP-2 and CD34+ cells are mentioned in tables 3.1.6 and 3.1.7 below.

Table 3.1.6: HUDEP-2 differentiation media composition.

Components	From	Stock Conc.	Working Conc.	For 10 ml	For 40 ml	For 500ml
IMDM glutamax	Gibco			10ml	40ml	500ml
AB serum	MP		3%	300ul	1.2ml	15ml
FBS	Gibco		2%	200ul	800ul	10ml
Insulin	Sigma		1ml/1000ml	10ul	40ul	500ul
Heparin 110U/mg	Sigma	1200U/ml	3U/ml	25ul	100ul	15mg
EPO	Zyrop	(2U/ul)	(3U/ul)	15ul	60ul	1500U
Holo-Transferrin	BB	80mg/ml	200ug/ml	25ul	100ul	100mg
SCF	Peprotech	100ng/ul	10ng/ml	2ul	8ul	100ul
IL3	Peprotech	5ng/ml	1ng/ml	2ul	8ul	100ul
Pen/Strep	Gibco	(100X)	(1X)	100ul	400ul	5ml
Doxycyclin	Sigma	(1mg/ml)	(1ug/ml)	10ul	40ul	1ml

Table 3.1.7: CD34+ HSC differentiation media.

Components	From	Stock Conc.	Working Conc.	For 10 ml	For 40 ml
IMDM glutamax	Gibco			10ml	40ml
AB serum	MP		5%	500ul	2ml
Insulin	Sigma		1ml/1000ml	10ul	40ul
Heparin 100U/mg	Sigma	1200U/ml	2U/ml	10ul	40ul
EPO	Zyrop	(2U/ul)	(3U/ul)	15ul	60ul
Holo-Transferrin	BB sol	330mg/ml	330ug/ml	20ul	80ul
SCF	Peprotech	100ng/ul	100ng/ml	10ul	40ul
IL3	Peprotech	5ng/ml	2.5ng/ml	5ul	20ul
Pen/Strep	Gibco	(100X)	(1X)	100ul	400ul
Hydrocortisone	Sigma		1uM	10ul	40ul

3.1.9 Analysis of Base editing efficiency:

The edited samples were used to isolate genomic DNA with a DNA isolation kit. For Sanger sequencing, PCR amplification was performed using GoTaq® Hot Start Polymerase premix. While for NGS, GXL premix was used to amplify the targets and sequenced using the MiSeq platform. The primers used for targeted amplification in case of Sanger sequencing and NGS are described in tables 3.1.1 and 3.1.8. Library preparation and sequencing for NGS were carried out following a previously described protocol from Corn Lab(Corn, 2017). The Fastq files were analyzed using CRISPResso-2 for evaluating base editing frequency(Clement, Rees, Matthew C Canver, *et al*., 2019). Sanger sequencing data was used to analyze indels and base editing efficiency using tools such as ICE and EditR respectively (Hsiau *et al*., 2018; Kluesner *et al*., 2018).

To analyze the editing in the *HBG1* and *HBG2* promoters individually, NGS 4F and NGS 2R primers were used. After NGS, the FastQ file was aligned to *HBG1* and *HBG2* promoter sequences, and editing efficiency was calculated using the ratio of

edited reads to total reads. The aligned read frequency was visualized using IGV (James T. Robinson *et al*, 2012; Langmead and Salzberg, 2013).

Table 3.1.8: Primer used to validate base editing in HBG promoter.

Sl.no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
Sanger sequencing primer for *HBG* promoter region				
1	HBF 1 F	acaaaagaagtcctggtatc	490bp	This study
2	HBF 2 F	ttactgcgctgaaactgtgg	772bp	This study
3	HBF 1 R	cttcccagggtttctcctcc		This study
NGS primers for HBG promoter region				
4	NGS 2 F	gctcttccgatct tgaatcggaacaaggcaaagg	325bp	This study
5	NGS 2 R	gctcttccgatct gtgaaatgacccatggcgtc		This study
6	NGS 3 F	gctcttccgatct cctggacctatgcctaaaaca	318bp	This study
7	NGS 3 R	gctcttccgatct agtttagccagggaccgttt		This study
8	NGS 4 F	gctcttccgatct cggctgacaaaagaagtcct		This study
Primers to differentiate HBG1 and HBG2				
9	HBG1 F	ccacagtacctgccaaagaa	940bp	This study
10	HBG2 F	ccatagtatctggtaaagagca	940bp	This study
11	HBG1/2 R	ggcgtctggactaggag		This study
Sanger sequencing primer for AAVS1 locus				
12	AAVS1 seqF	gagatggctccaggaaatgg	420bp	Matthew J.J *et al*., Sci Rep,2018.
13	AAVS1 seqR	acctctcactcctttcatttgg		Matthew J.J *et al*., Sci Rep,2018.
14	CCR5 seqF	cagctgcttggcctgttagt	413 bp	This study
15	CCR5 seqR	tccctcttgtctggaggaaa		This study

3.1.10 Real-time PCR:

The cells that were edited had their total RNA isolated using the NucleoSpin RNA kit from Macherey-Nagel. This RNA was then reverse transcribed with the cDNA Synthesis Kit from iScript™ Bio-Rad. The *HBB*, *HBA*, and *HBG* genes' relative expression (ΔΔCT) was measured using qRT-PCR with SsoFast™ EvaGreen® Supermixes from Bio-Rad on a QuantStudio™ 6 Flex Real-Time PCR System from Applied Biosystems. A 10 µl qRT-PCR mixture containing 1 µl each of the forward and reverse primers (5 µM) and 1 µl of 5-fold diluted cDNA template, among other ki components were used. To normalize the data for ΔΔCT, *GAPDH* was used as an internal control gene. The manufacturer's protocol was followed for the cycling condition, and dissociation curve analysis was performed to ensure no unspecific amplification occurred.

The vector copy number (VCN) was evaluated using qRT-PCR on genomic DNA from the transduced samples, similar to the method described in Barczak *et al*, with some modifications. The U6 promoter was targeted for gRNA integration, the Cas9 gene was targeted for Cas9 and base editors' integration, and by target WPRE both gRNA and Cas9 integration were evaluated. Primer targeting Exon 2 of the *HBB* gene was used as a single copy gene-specific reference. The standards used were pLKO5.sgRNA. EFS.GFP (Plasmid #57822, Addgene) and an in-house plasmid with *HBB* CDS (details not provided). The primers used for evaluating the relative expression and to determine the VCN were provided in table 3.1.9.

Table 3.1.9: Primers used for qRT PCR.

S. No	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
qRT PCR primers to amplify globin gene expression				
1	HBB E2E3 RT F	acctttgccacactgagtgag	110bp	This study
2	HBB E2E3 RT R	tttgccaaagtgatgggcca		This study
3	HBA RT F	cgacaagaccaacgtcaagg	99bp	This study
4	HBA RT R	gtggggaaggacaggaacat		This study
5	HBF E2E3 RT F	cttccttgggagatgccata	136bp	This study
6	HBF E2E3 RT R	aaaacggtcaccagcacatt		This study
7	GAPDH E7E8 RT F	ctgcaccaccaactgcttag	110bp	This study
8	GAPDH E7E8 RT R	gtcttctgggtggcagtgat		This study
qRT PCR primers to validate Vector copy number				
9	Cas 9 var RT F	ccgaagaggtcgtgaagaag	128bp	This study
10	Cas 9 var RT R	gccttatccagttcgctcag		This study
11	U6 RT F	agggcctatttcccatgatt	151bp	This study
12	U6 RT R	aaactgcaaactacccaagaaa		This study
13	WPRE RT F	caccacctgtcagctccttt	135bp	This study
14	WPRE RT R	acaacaccacggaattgtca		This study
15	HBB locus ctrl F	ttggacccagaggttctttg	123bp	This study
16	HBB locus ctrl F	gagccaggccatcactaaag		This study

3.1.11 HbF intracellular staining:

Edited cells were spun at 1000 rpm for 5 min at RT. The pellet was resuspended in 200µl BPBS (containing 1xPBS and 0.1% BSA) and transferred to a 96 well cell culture plate. The plate was spun at 1000rpm for 5min, and then the pellets were resuspended in freshly prepared 100µl of 0.05% glutaraldehyde (MP Biomedicals) using BPBS and incubated at RT for 8 min. Each well was topped up with 100 µl of BPBS, spun at 1000 rpm for 5 min, and the pellet was washed with 200µl BPBS. 100µl of 0.1% fresh Triton x-100 (Fisher scientific) prepared in BPBS was then added to the pellet, resuspended, and incubated at RT for 7 min. After 7 minutes, 100µl of BPBS was added,

centrifuged at 1000 rpm for 5min, and the pellet was again washed with 200 µl BPBS. 50µl of BPBS containing 2µl of HbF antibody (Fetal Hemoglobin Monoclonal Antibody (HbF-1), APC - Thermo Fisher) was added to each well except the unstained and kept in the dark for 15min. After topping it up with 100µl of BPBS, the plate was spun at 1000 rpm for 5min, the supernatant was discarded, and the pellet was washed again with 200µl BPBS. The pellet was resuspended in 300 µl of BPBS and taken for FACS analysis (BD FACS ARIA-III or CytoFLEX LX Flow Cytometer - BC) to check for HbF positive cells based on APC fluorochrome staining.

3.1.12 Detection by High Performance Liquid Chromatography (HPLC):

The differentiated cells were harvested, washed with 1x PBS, and suspended in 1100 µl of cold ddH20. After sonication for 30 seconds with 50% amplitude on ice, the cells were centrifuged at 14000 rpm for 15 minutes at 4°C. The resulting supernatant (1000 µl) was analyzed for hemoglobin variants using the VARIANT II Hemoglobin Testing System from Bio-Rad. The hemoglobin percentages were determined using Bio-Rad's Clinical Data Management (CDMTM) Software.

The remaining 100 µl of the supernatant was subjected to reverse phase HPLC using Shimadzu Corporation-Phenomenex. Globin chains were measured using a Shimadzu UFLC consisting of binary gradient pumps (LC-20A), autosampler (SIL-HTC), and a column oven (CTO-20AC) coupled with UV detection (ShimadzuTM, Kyoto, Japan). The data was analyzed using LC SolutionsTM software. Chromatographic separation of the analytes was done using Aeris Widepore 3.6 lm XB-C18 25cm 4.6mm column behind a SecurityGuard UHPLC Widepore C18 4.6mm guard column (PhenomenexTM) (Loucari *et al*., 2018). The LC conditions were as follows: Solvent A: 0.1% Trifluoroacetic acid (TFA), pH 3.0 and Solvent B: 0.1% TFA in acetonitrile was used as mobile phase with gradient elution (40% Solvent B) at a flow rate of

1.0ml/min and column temperature maintained at 70 C. The total run time is 8 minutes. The UV detection was set at 190nm for globin chain detection.

3.1.13 Validation of 4.9kb large deletion:

To measure the extent of 4.9kb large deletion in the *HBG* promoter region, we conducted qPCR as described previously by Li *et al* (Li *et al*., 2018). The primers used are mentioned in table 3.1.10. Additionally, to confirm the impact of the larger deletion on the expression of gamma-globin, we performed a globin chain analysis on the differentiated erythroid cells using RP HPLC. The percentage of A gamma and G gamma-globin chains obtained from each sample was normalized to the control.

Table 3.1.10: qRT PCR primers to validate 4.9kb large deletion.

S. No	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	Del_qPCR_F	aggggctcaacgaagaaaaagtgt		Chang Li *et al*., Blood, 2018.
2	Del_qPCR_R	cacttcattgtagttaccgtggaaaga		Chang Li *et al*., Blood, 2018.
3	Del_ctrl_qPCR_F	aaatgaatcagcagaggctcac		Chang Li *et al*., Blood, 2018.
4	Del_ctrl_qPCR_R	atgcactaacatccaactatacaaaa		Chang Li *et al*., Blood, 2018.

3.1.14 COS-7 cell transfections and nuclear extraction:

COS-7 cells were transfected with 5 µg of a mammalian expression plasmid, either pcDNA3-empty (Invitrogen) or pSG5/mEKLF-Mouse (Miller and Bieker, 1993), using FuGENE® 6 (Promega) in a 10 cm culture dish following the manufacturer's guidelines. The cells were incubated at 37°C for 48 hours before being harvested. Nuclear extractions were performed in accordance with the procedure described by Andrews and Faller (Miller and Bieker, 1993).

3.1.15 Electrophoretic Mobility Shift Assay (EMSA):

The P-32 label was attached to the sense strand of each probe using T4 PNK (NEB) and γ-32P ATP (Perkin Elmer), followed by annealing with the antisense strand through gradual cooling from 100°C to room temperature. The radiolabelled probes were purified by quick spin columns for DNA purification (Roche). Plasmids were overexpressed and extract devoid of the target protein was utilized to assist in identifying the background bands caused by endogenous protein binding. The KLF1 antibody was employed to detect the protein on the gel (Crossley *et al*., 1996). The complexed samples were loaded onto a 6% native polyacrylamide gel in TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA), and electrophoresis was conducted at 4°C for 1 hour and 40 minutes at 250 V. Subsequently, vacuum drying was done before exposing the FUJIFILM BAS Cassette2 phosphor screen overnight, and imaging was performed on a GE Typhoon FLA 9500 fluorescent image analyzer. The oligonucleotides utilized in the radiolabeled probes are detailed in table 3.1.11.

Table 3.1.11: Probes used for EMSA.

Sl.no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	HBG2 -110 to -132 Wild type (WT)	ggccagccttgccttgaccaata	Sense	This study
2	HBG2 -110 to -132 Wild type (WT)	tattggtcaaggcaaggctggcc	Antisense	This study
3	HBG2 -110 to -132 -123T>C/-124T>C	ggccagcccgccttgaccaata	Sense	This study
4	HBG2 -110 to -132 -123T>C/-124T>C	tattggtcaaggcgggctggcc	Antisense	This study
5	Hbbt1 CACCC (KLF1 +ve control)	tagagccacaccctggtaag	Sense	Merlin C *et al*., Mol Cell Biol, 1996.
6	*Hbbt1 CACCC (KLF1 +ve control)*	cttaccagggtgtggctcta	Antisense	Merlin C *et al*., Mol Cell Biol, 1996.

3.1.16 ChIP qPCR

The immunoprecipitation procedure was carried out using 5×10^7 cells of both wild type and edited HUDPE2 cells prior to their differentiation. The cells were cross-linked using 1% formaldehyde solution for 10 minutes at room temperature and then treated with glycine to a final concentration of 125 mM to quench the reaction. The cross-linked cells were lysed and sonicated to generate chromatin fragments of approximately 200-300 bp over 10 cycles of 30 seconds each with 30 seconds intermissions at 4°C. For immunoprecipitation, 100 µl of Dynabeads™ Protein G complexed with either 15 µg of KLF1 antibody or normal rabbit IgG was added to the chromatin fragments and incubated overnight at 4°C. The magnetic beads were separated and washed thoroughly before the elution step, and crosslinking was reversed by incubation at 65°C overnight. The DNA was then purified, and its quantity was determined using SYBR green reagents and the ΔΔCt method with reference to whole cell extract on a ViiA 7 Real-Time PCR System for specific targets (Table 3.1.12).

Table 3.1.12: qRT PCR primers used for ChIP analysis.

Sl.no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	*γ-globin* promoter -73 to -179 F	caaatatctgtctgaaacggtccc	106 bp	Gabriella E. M *et al*., Blood, 2019.
2	*γ-globin* promoter -73 to -179 R	actctaagactattggtcaagtttgc		Gabriella E. M *et al*., Blood, 2019.
3	*γ-globin* promoter -199 to -106 F	tcaatgcaaatatctgtctgaaacg	93bp	Gabriella E. M *et al*., Blood, 2019.
4	*γ-globin* promoter -199 to -106 R	caaggctattggtcaaggcaa		Gabriella E. M *et al*., Blood, 2019.
5	*SP1* promoter F	acctctccgcccactagga		Beeke Wienert *et al*., Blood, 2017.
6	*SP1* promoter R	caacggccaaccagaatcc		Beeke Wienert *et al*., Blood, 2017.
7	*VegfA* F	ggtttgtatcctgcccttc		Beeke Wienert *et al*., Blood, 2017.
8	*VegfA* R	actgggtcttgctgttttcc		Beeke Wienert *et al*., Blood, 2017.

3.1.17 RNA sequencing and analysis:

To assess the quality of extracted total RNA, the NucleoSpin RNA kit (Macherey-Nagel) was used and analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies). Polyadenylated transcripts were purified from 1 μg of total RNA using oligo-dT beads (TruSeq RNA Sample Preparation Kit, Illumina) and fragmented in the presence of divalent cation. Reverse transcription was carried out using the Superscript II Reverse Transcriptase kit (Life Technologies) and cDNA was purified using Ampure XP SPRI beads (Beckman Coulter) before Illumina adapter ligation and amplification. The quality and quantity of libraries were evaluated using a NanoDrop spectrophotometer (Thermo Scientific) and Bioanalyzer (Agilent Technologies), respectively. The Illumina NovoSeq 6000 platform generated 150-bp paired-end reads, which were converted to FastQ files using bcl2fastq and trimmed with TrimGalore to remove low-quality bases, adapter sequences, and unpaired sequences. Trimmed reads were aligned to the Homo sapiens genome assembly GRCh38. NFCore RNA Seq pipeline was used to analyze expressed transcripts quantitatively and qualitatively (Ewels *et al*., 2020). The data can be accessed through the GEO Series accession number GSE192801.

Transcriptome analysis was performed in duplicate using wild type HUDPE2, ABE and CBE stable cells with or without gRNA 2 and 11. Gene expression levels were normalized and differential expression analysis was carried out using the Interactive Gene Expression Analysis Kit (iGEAK) RNA-seq v1.0, an R and JavaScript-based tool. The transcript was counted from the sorted bam files by the aligner mentioned above (Ewels *et al*., 2020).

3.1.18 Off-target analysis

To identify Cas-dependent DNA off targets, Cas-OFFinder was utilized, with a maximum of 3 mismatches allowed for target selection. The targets were then amplified

and sequenced using the Illumina MiSeq platform. CRISPResso2 was employed to align reads, but only high-quality reads with q=30 were used for analysis. The targets identified and the primer used for targeted amplification are provided below (Tables 3.1.13 and 3.1.14). REDItools v 2 was utilized to calculate transcriptome-wide A-to-I and C-to-U conversion in ABE and CBE edited samples. All nucleotides except the respective nucleotide (A for ABE and C for CBE) were excluded from the analysis. Read coverage and quality criteria were in accordance with previous studies, Koblan *et al* (Koblan, Erdos, *et al*., 2021). The frequency of A converted to I/N and C converted to U/N were calculated by dividing the total number of converted nucleotides by the respective nucleotides after filtering (A-to-(I or N)/A*100 or C-to-(U or N)/C*100). The experiment was carried out as two biological replicates.

Table 3.1.13: The targets analyzed for DNA off-target

Sl. no	Name	DNA offtarget target seq with PAM	Chromosome	Position	Mismatches
1	OT1	ttgggttgcccagacttgccagg	chr1	25039865	3
2	OT2	gtggggtggccagcctttcctgg	chr1	235427633	3
3	OT3	gtggggtggccagccttccctgg	chr2	64808323	3
4	OT4	atggggtgtgcagccttgcctgg	chr12	113177162	3
5	OT5	aggggttggcccgccttcccagg	chr16	674938	3
6	OT6	ttggcttggccagacttgcccgg	chr9	110575343	3
7	OT7	gtggggtggccagccttcccagg	chr9	118816545	3
8	OT8	atgggttgtccagcgttgcttgg	chrX	150946920	3
9	OT9	atgggttggccaggctggtctgg	chr6	7187976	3
10	OT10	gtggggtggccagccttccctgg	chr10	119547882	3
11	OT11	atgggttgaccaacctggccagg	chr13	86143790	3

Table 3.1.14: Primers to amplify DNA off-target

Sl. no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	G11 OT1 F	tacacgacgctcttccgatct ggcagaggggacacatcagt	192bp	This study
2	G11 OT1 R	agacgtgtgctcttccgatct tgcgctatgatgttgggtatgtcc		This study
3	G11 OT2 F	tacacgacgctcttccgatct acttaatgctcacagattggttcaatcaggc	190bp	This study
4	G11 OT2 R	agacgtgtgctcttccgatct agccaatcaggttcaagatggcag		This study
5	G11 OT3 F	tacacgacgctcttccgatct gtctactggaaagcccatttgcatagtaag	200bp	This study
6	G11 OT3 R	agacgtgtgctcttccgatct gtgagagagagaggacttctgag		This study
7	G11 OT4 F	tacacgacgctcttccgatct tcctcagttccaagccttggg	194bp	This study
8	G11 OT4 R	agacgtgtgctcttccgatct gtcatccccaatccacaaactcac		This study
9	G11 OT5 F	tacacgacgctcttccgatct cacacgtgtcttatctgtcacctc	193bp	This study
10	G11 OT5 R	agacgtgtgctcttccgatct ctgaccactcttgcaactccatg		This study
11	G11 OT6 F	tacacgacgctcttccgatct atcagggaaaaacctgccatgtgc	191bp	This study
12	G11 OT6 R	agacgtgtgctcttccgatct gcaattcctctgttcaagacaggatgtag		This study
13	G11 OT7 F	tacacgacgctcttccgatct gggcaaggggaaacttcatttgca	192bp	This study
14	G11 OT7 R	agacgtgtgctcttccgatct gagaggcgacttctgagagg		This study
15	G11 OT8 F	tacacgacgctcttccgatct ttacaggcatgaaccaccacatcc	194bp	This study
16	G11 OT8 R	agacgtgtgctcttccgatct gtcagagaactgagcctaatggag		This study
17	G11 OT9 F	tacacgacgctcttccgatct cacttgagatcaggatttggagaccag	192bp	This study
18	G11 OT9 R	agacgtgtgctcttccgatct gcgcaatcttgggtcactgca		This study
19	G11 OT10 F	tacacgacgctcttccgatct ggcccacagattggttagatcag	194bp	This study
20	G11 OT10 R	agacgtgtgctcttccgatct ctacttgggactaggcatgttcag		This study
21	G11 OT11 F	tacacgacgctcttccgatct gggggctactctaataaaaagactttaggaac	200bp	This study
22	G11 OT11 R	agacgtgtgctcttccgatct gttccttatcagcttatgtagattttgggctgag		This study

3.1.19 Animal Studies:

The CD34+ cells, which were subjected to nucleofection with base editing mRNA, were transplanted into NBSGW mice three hours after editing (McIntosh *et al*., 2015). Pretransplant conditioning was performed 48 hours prior to the infusion of edited cells in 6-8-week-old mice, using intraperitoneal administration of busulfan (12.5 mg/kg) (Ref). The edited cells were suspended in 100 µl of 1xPBS and administered via tail vein injection. After a duration of 15 weeks, the mice were euthanized, and the engraftment potential was assessed by validating the ratio of hCD45+ cells to the total number of hCD45+ and mCD45+ cells, multiplied by 100. Bone marrow (BM) (Amend *et al*., 2016), spleen, and peripheral blood samples were collected at week 15, and analyzed for hCD45+ (FITC) (BD) cells and mCD45+ cells (APC) (BD) using flow cytometry (CytoFlex-BC). For peripheral blood analysis, red blood cells (RBC) were removed by treating the samples with RBC lysis buffer before processing with antibodies for flow cytometry analysis. BM cells were also fixed for intracellular staining with HbF (APC) along with hCD235 (PE) to evaluate in vivo HbF regulation in the erythroid progenitor cells (CD235+ cells). DNA was isolated from BM using the NucleoSpin Blood kit (Macherey-Nagel) according to the manufacturer's instructions, and the HBG promoter region was amplified using PCR. The amplified product was subjected to Sanger sequencing to evaluate the efficiency of editing.

3.1.20 Statistics:

Statistical analyses were conducted using GraphPad Prism 8.1. Normal distribution of all data allowed for the use of appropriate statistical tests, such as unpaired two-sided t-tests or one-way ANOVA. Statistical significance was considered at $p<0.05$. Linear regression analysis was performed to determine any correlations between two variables, and Pearson correlation was used to identify relationships between samples. Finally, PCA was performed using the R statistical package.

Section 3.2: Monoallelic integration of Factor IX transgene in ITGA2B gene and their concerted expression can ameliorate haemophilia B

3.2.1 gRNA designing and cloning:

The guide RNAs (gRNAs) were designed using an online tool called "CHOPCHOP (version 3)", which allows for the design of efficient gRNAs for the CRISPR gene editing system (Labun *et al*., 2019). The efficiency of the top 10 gRNAs obtained from CHOPCHOP and another tool called Microhomology-Predictor (CRISPR RGEN tools) were compared (Bae, Kweon, *et al*., 2014). The details of the gRNAs are provided below (Table 3.2.1). To select the most promising gRNA for further validation, a dimension reduction analysis was carried out. The selected gRNAs were synthesized and cloned into a lentiviral vector called pLKO5.sgRNA.EFS.GFP/RFP vector, available from Addgene (#57822) (Stone *et al*., 2014). The process of cloning, PCR amplification, and confirmation of DNA sequence using Sanger sequencing was performed in a manner as described in the study-I.

Table 3.2.1: The gRNAs used in this study to screen ITGA2B Intron-1.

Sl.no	Names	gRNA Info 5'-3'	Efficiency (CHOP CHOP)	MMEJ Score (Rgen)
1	gRNA 1	TAGCGGGTACCTCTAAAGTGGGG	69	7036
2	gRNA 2	AATACGAGTAACGGTTAGCATGG	60	4156
3	gRNA 3	TGATTGCCCTCGCTGAGAGTCGG	55	4828
4	gRNA 4	GGTAGCGGGTACCTCTAAAGTGG	49	7547
5	gRNA 5	GCTCAACTGCTATCGCAAATGGG	44	4747
6	gRNA 6	GAACCGCTGGTTCTTGTTGCGGG	38	7124

7	gRNA 7	TCTCAGCGAGGGCAATCACGTGG	67	4825
8	gRNA 8	GCGTTTGTTATGTATTTGGCTGG	44	3992
9	gRNA 9	GTGAAACCGACTCTCAGCGAGGG	66	5016
10	gRNA 10	GATCTGGTAACAGTCTAGGCAGG	60	5819

3.2.2 Donor constructs:

In this study, the AAV plasmids, pAAV-MCS (transfer plasmid), and pHelper-AAV (helper plasmid) were acquired from Agilent technologies (AAV Helper-Free System), while the AAV6 capsid (rep/cap plasmid) was obtained as a gift from Sanjay Kumar and Arun Srivastava. Four donor constructs were generated in the transfer plasmid backbone, all designed with common 500bp homology arms. These donor constructs include: 1) AAV6-Ef1α -GFP, 2) AAV6-SA-P2A-GFP, 3) AAV6-IRES-GFP, and 4) AAV6-IRES-F9. For the construction of AAV6-EF1A -GFP, the entire donor template, including the homology arms, was synthesized as a gene fragment by GENEWIZ and cloned into the transfer plasmid using the NEB HIFI assembly kit, following the manufacturer's instructions. To construct promoterless donors like AAV6-SA-P2A-GFP and AAV6-IRES-GFP, the backbone obtained from AAV6-EF1A -GFP, digested with Xba1 and Mlu1, was used. For cloning AAV6-SA-P2A-GFP, the splice acceptor (SA) was amplified from the endogenous ITGA2B gene (100bp of intron 1 and 61bp exon 2) with the right primer having the sequence of P2A (self-cleaving peptide). For cloning IRES in the AAV6-IRES-GFP vector, the EMCV-IRES was synthesized as a gene fragment from IDT. The GFP fragment without a promoter, along with the bGH poly (A) signal required for both promoterless donors, was amplified using PCR. Finally, the respective fragments, along with the digested AAV6-EF1A -GFP backbone, were cloned using the NEB HIFI Assembly, as per the kit's instructions, where the

homology arm length varied from 25 to 30 bp. To construct AAV6-IRES-F9, the exact protocol for cloning AAVRNS8.3 was followed, but instead of the amplified GFP fragment, a FIX gene fragment synthesized from GENEWIZ was used for cloning. The primers used for cloning donor constructs were mentioned in table 3.2.2 below. The ITR integrative is confirmed with restriction digest with SmaI and AhdI enzymes (NEB) (Bak *et al*., 2018).

In addition, the pLKO5.sgRNA.EFS.GFP vector (Addgene #57822) was employed to clone gRNA, and the Cas9 vector lentiCRISPR V2.1, which was previously generated by our laboratory, was also used (Ravi *et al*., 2022). The selection criteria for positive clones were carried out as described in study 1. To generate lentivirus, the pMD2.G and psPAX2 lentiviral constructs (Addgene #12259, 12260) - both second-generation lentiviral packaging constructs - were utilized.

Table 3.2.2: Primers used for cloning.

Sl.no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	SA2A F	aaaaccagctctcaagaggggatctggtaatctagacc ccatgccacatggccttccgggcagtgc	293 bp	This study
2	SA2A R	aacagctcctcgcccttgctcaccatggatccaggtcc agggttctcctccacgtctccagcctgcttcagcaggct gaagttagtagcctcgaggccgcccgtctcctcctggc tggg		This study
3	GFP cloning F	atggtgagcaagggcgaggagctgttcacc	1001 b p	This study
4	GFP cloning R	aacaagaaccagcggttcacatgctccctggaatgcct acgcgtccatagagcccaccgcatccccagca		This study

3.2.3 Cell culture:

The HEL and HEK 293T cell lines were grown in RPMI (Roswell Park Memorial Institute) media and DMEM (Dulbecco's Modified Eagle's Medium), respectively, and supplemented with 10% fetal bovine serum (FBS, Gibco) and 1x penicillin-streptomycin (PS, Gibco). The culture conditions and components for CD34+ cells were kept consistent with those described in Study-I.

3.2.4 Virus production:

To assess the effectiveness of gRNA, lentivirus were produced and transduced in the HEL cell line using methods outlined in Study-I. In order to produce AAV6 virus, HEK293T cells were cultured in a 10 cm cell culture dish until reaching 80%-90% confluency. The cells were then transfected with 4 µg of AAV6 capsid, 6 µg of pHelper-AAV, and 3 µg of the transfer plasmid containing the donor template using PEI (1mg/ml) (Polyscience MW 250,000). A complex consisting of 450 ul of optiMEM and 78 ul of PEI (6ul for 1ug plasmid) was added to the plasmids, mixed well by vortexing, and incubated for 20 min. The lipoplex was then added to the plate, mixed gently, and incubated in a 37°C incubator with 5% CO2 for 12 hours. After 12 hours, the media was removed, and the adhered cells were washed twice with 10 ml of 1x PBS to remove un-transfected lipoplex and plasmids. The cells were then resuspended with 10 ml of complete DMEM. The viral supernatant was collected at 72 hrs and 144 hrs and clarified by centrifuging at 3000rpm. The clarified supernatant was then ultracentrifuged at 36000rpm for 2 hours at 4°C. After pelleting the viral particles, the supernatant was removed, and the pellet was resuspended in 200ul of PBS with 0.001% pluronic F68 and stored at -80°C (Bak *et al*., 2018; Piras *et al*., 2016).

3.2.5 AAV donor vector titration:

To extract the viral genome, 10ul of AAV is mixed with 40ul of QuickExtract DNA extraction solution and incubated in a thermal cycler for 6 minutes at 65°C followed by 3 minutes at 98°C (Bak *et al*., 2018). Four different dilutions of the DNA are prepared: 200, 400, 800, and 1600-fold. The dilutions are then subjected to qRT-PCR using standards ranging from 1E2-1E8 copies/ul, as well as a no-template control (NTC). The donor plasmid used for virus production is used as a standard, and the assay is performed in replicates. The primers used are listed in table 3.2.3 below. The viral concentration is measured manually using the standard values obtained, and the average value of all fold dilutions performed is used to calculate the final multiplicity of infection (MOI).

Table 3.2.3: qRT PCR primers to validate Multiplicity of infection (MOI).

Sl. no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	MOI GFP F	gacgtaaacggccacaagtt	78 bp	This study
2	MOI GFP R	gaacttcagggtcagcttgc		This study
3	MOI F9 F	acattgcccttctggaactg	129 bp	This study
4	MOI F9 R	tccccagccacttacatagc		This study

3.2.6 Virus transduction:

For lentivirus transduction, the same protocol was used as described in Study-I. In case of AAV-based transduction, cells were incubated with the titrated AAV for 12 hours, after which the old media was replaced with fresh media.

3.2.7 Nucleofection:

To prepare for HDR, 1 million target cells were pelleted and suspended in 13 µl of Lonza P3 solution, which was prepared as recommended by the manufacturer. A RNP complex consisting of Cas9 (5 µl, 100 pmole from Takara) and gRNA (1 µl, 100 pmole from synthego) was added to the cells, and the mixture was incubated at room temperature for 10 minutes. The volume was adjusted to 20 µl with P3 solution, and the mixture was added to a single well of the 16-well nucleofection strip. The cells were nucleofected with a 4D-Nucleofector® X Unit (Lonza) using the pulse code DZ100. After nucleofection, the cells were incubated in the strip at room temperature for 5 minutes. The cells were then gently aspirated and resuspended in complete media.

3.2.8 Polymerase chain reaction (PCR):

The Whole Length PCR (WLP) involves designing primers that amplify the integrated fragment and the wild-type fragment, without amplifying the donor construct, by placing the primers beyond both the homology flanks. On the other hand, in the case of In Out PCR (IOP), only the HDR sample and not the control is amplified by designing the left primer beyond the homology flanks while the right primer is designed in the integrated fragment. The PCR's were carried out using GXL premix (Takara) following the manufacturer's instructions. The extension time was 3 minutes and 1 minute for WLP and IOP, respectively. The table 3.2.4 provides the details of the primers used.

Table 3.2.4: Primers to detect HDR.

Sl.no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	WLP F	gcctgaatcagctcaactcc	1382 + insert	This study
2	WLP R	attgttccctgtgccctgta		This study

3	IOP GFP F	gcctgaatcagctcaactcc	Varies by donor	This study
4	IOP GFP R	aagtcgtgctgcttcatgtg		This study
5	IOP IRES F9 F	tgaagggtcattctagttccaaag	877 bp	This study
6	IOP IRES F9 R	cacggcgactactgcactta		This study

3.2.9 Digital droplet Polymerase chain reaction (ddPCR):

To confirm and quantify integration, a probe-based multiplexing ddPCR was performed. For target-primer design, the left primer and the probe were designed beyond the homology flank, while the right primer was designed in the integrated sequence. A reference-primer was designed in Exon-1 Intron-1 junction of the ITGA2B gene, the primer and probe information are provide below (Table 3.2.5). Genomic DNA was isolated from the samples using MN Blood DNA kit, following the manufacturer's protocol. The isolated DNA samples were then subjected to restriction digestion, where a 20 µl reaction containing 250ng of DNA, along with an appropriate restriction enzyme (NEB), was incubated for 1 hour at 37°C, followed by heat inactivation at 80°C. For donor template AAV6-EF1α and AAV6-SA-P2A EcoRV was used, and for AAV6-IRES BamH1 was used to set up restriction digestion. The reaction mixture contained 4x ddPCR Multiplex Supermix (BioRad), forward and reverse primers for both the target and reference (1 pmol working each), along with the respective probes (0.25 pmol working each), and 25 ng of digested product in a 25 µl reaction. Droplets were generated using a droplet generator, transferred to a 96-well plate, heat-sealed, and PCR was carried out as per the manufacturer's instructions (QX200 Droplet Digital PCR System). The PCR cycling conditions included an initial denaturation at 94°C for 10 min, 35 cycles of 94°C for 30 s, 60°C for 45 s, 72°C for 3 min with 1°C/sec ramp,

followed by a final denaturation at 98°C for 5 min, and hold at 4°C. Finally, the plate was read using a plate reader, and the percentage of HDR frequency was calculated by dividing the target concentration of droplets by the reference concentration of droplets.

Table 3.2.5: ddPCR primers and probes to detect knock-in efficiency

Sl. no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	EF1α F	tgaagggtcattctagttccaaag	1010 bp	This study
2	EF1α R	cacggcgactactgcactta		This study
3	P2A F	tgaagggtcattctagttccaaag	910 bp	This study
4	P2A R	cagcaggctgaagttagtagc		This study
5	IRES F	tgaagggtcattctagttccaaag	877 bp	This study
6	IRES R	taggaatgctcgtcaagaagac		This study
7	Locus F	aagggaagttggggtattgg	995 bp	This study
8	Locus R	gacatgtgctcctctccaca		This study
9	Target Probe	atcagctcaactccttcctgtgcc		This study
10	Locus Probe	aaggtgttgagggtggagtggttc		This study

3.2.10 Megakaryopoiesis:

CD34+ cells were subjected to megakaryocyte differentiation using two-phase media. Phase-I media, consisting of SCF, Flt3, IL6, TPO, and PS (1x), was used for the initial 24 hours of culture. After this time, the media was changed to phase-II media, containing SCF, IL6, TPO, and PS (1x), which was replaced every other day from day 4 to day 10. The cells were maintained at a seeding density of 0.5 million cells/ml, the details of the megakaryocytic differentiation media components are provided below (Table 3.2.6 and 3.2.7) (Matsunaga *et al*., 2006; Perdomo *et al*., 2017; Zarif *et al*., 2011).

From day 4-10, the cells were analyzed for megakaryopoiesis by flow cytometry using CD 41a (APC) and CD 42b (PE) markers. Ploidy analysis was also performed on day 10 using 0.3 million cells. The cells were fixed, permeabilized with glutaraldehyde and Triton, and resuspended in 50 µl PBS. Then, q µl DAPI (Sigma, 1 mg/ml) and 1 µl CD41 APC (BD) were added and mixed well, followed by a 5-minute incubation. Next, 0.5ul of 10mg/ml RNAse A (Novogen 5ug/rn) was added, and the cells were incubated at room temperature for 10 min, pelleted, and washed with PBS twice. Finally, the cells were resuspended in 100ul PBS, and the intensity of DAPI was measured by flow cytometry in 50,000 CD41a high events.

Table 3.2.6: Phase I Megakaryocytic differentiation media composition.

Components	From	Stock Conc.	Working Conc.	For 10 ml	For 40 ml
StemSpan™ SFEM II	STEMCELL Technologies			10ml	40ml
SCF	Peprotech	(100ng/ul)	(10ng/ml)	1ul	5ul
FLT3	Peprotech	(100ng/ul)	(10ng/ml)	1ul	5ul
IL6	Peprotech	(100ng/ul)	(10ng/ml)	1ul	5ul
TPO	Peprotech	(50ng/ul)	(50ng/ml)	10ul	40ul
Pen/Strep	Gibco	(100X)	(1X)	100ul	400ul

Table 3.2.7: Phase II Megakaryocytic differentiation media composition.

Components	From	Stock Conc.	Working Conc.	For 10 ml	For 40 ml
StemSpan™ SFEM II	STEMCELL Technologies			10ml	40ml
SCF	Peprotech	(100ng/ul)	(1ng/ml)	0.5ul	2ul
IL6	Peprotech	(100ng/ul)	(1ng/ml)	0.5ul	2ul
TPO	Peprotech	(50ng/ul)	(50ng/ml)	10ul	40ul
Pen/Strep	Gibco	(100X)	(1X)	100ul	400ul

3.2.11 Platelet:

On day 10 of megakaryopoiesis, approximately 0.1 million cells were extracted and subjected to flow cytometry analysis for platelet markers CD41a (APC) and CD42a (BV421). Once the presence of platelets was confirmed, platelet activation was induced by treating platelets with an agonist such as Thrombin Receptor Activating Peptide (TRAP). To activate platelets, 0.3 million cells were treated with 25 μM TRAP (Sigma Aldrich) and incubated for 10 minutes. CD41a (APC) and CD42a (BV421) was added to the mixture to gate platelets, and CD62p (PE Cy5) was added to estimate the number of activated platelets. The positive population of CD62p was analyzed from the CD41a+ and CD42a+ population.

To determine the number of platelets in the culture, the cells were gently mixed with a 1ml pipette, and 100ul of the culture was extracted and stained with 2 μL of CD42a Bv421 antibody for 10 minutes at 4 °C in the dark. The CD42a+ platelets in the 100ul were then counted using the counting mode in BC CytoFlex. Finally, the number of platelets per ml was extrapolated.

3.2.12 Co-localization assay:

Cytospin was performed using 50,000 HDR edited HEL cells at 400rpm for 10 minutes. Subsequently, the cells were treated with ice-cold methanol for 10 minutes, followed by a PBS wash. For antibody staining, CD62p (PE Cy5) antibody diluted in PBS (1ul in 50ul PBS) was added, and the cells were incubated in the dark at 4°C for 10 minutes. After washing the slide with PBS, the cells were rehydrated in PBS for 30 minutes at 4°C. Once complete, the slide was air-dried and visualized under a confocal microscope using a 100x objective (oil immersion) with the appropriate lasers.

3.2.13 Quantitative real time polymerase chain reaction (qRT PCR):

Total RNA was extracted from the cells using the NucleoSpin RNA kit (Macherey-Nagel) following the manufacturer's protocol. The extracted RNA was then reverse transcribed using a cDNA synthesis kit (iScript Bio-Rad). To evaluate the relative expression levels of GFP, FIX, and ITGA2B, qRT-PCR was performed using TB Green Premix Ex Taq II (Tli RNase H Plus) (Takara) as per the manufacturer's instructions in a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). The internal gene control GAPDH was used for ΔΔCT analysis. The respective primers used are mentioned in the table 3.2.8.

Table 3.2.8: qRT PCR primers to amplify globin gene expression.

Sl. no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	ITGA2B RT Exon1_F	gagctttgtgtccactgcaa	104 bp	This study
2	ITGA2B RT Exon1_R	actgggtccaggttcaagg		This study
3	ITGA2B RT Exon30_F	gggccagcaaatcatctgta	108 bp	This study
4	ITGA2B RT Exon30_R	tccccctcttcatcatcttc		This study
5	GFP RT F	gacgtaaacggccacaagtt	78 bp	This study
6	GFP RT R	gaacttcagggtcagcttgc		This study
7	Factor IX_RT_E1/E2_F	cctcatcaccatctgccttt	102 bp	This study
8	Factor IX_RT_E1/E2_R	tacctctttggccgattcag		This study
9	GAPDH E7E8 RT F	ctgcaccaccaactgcttag	109 bp	Ravi NS *et al*., elife, 2022.
10	GAPDH E7E8 RT R	gtcttctgggtggcagtgat		Ravi NS *et al*., elife, 2022.

3.2.14 Off-target analysis:

The potential effects of Cas-dependent DNA off-target editing were predicted and analyzed according to the methodology outlined in Study-1. The predicted off-targets and the primer used to target sequence the regions are provided in table 3.2.9 and 3.2.10 (Bae, Park, *et al*., 2014; Clement, Rees, Matthew C. Canver, *et al*., 2019).

Table 3.2.9: The targets analyzed for Cas dependent DNA off-target

Sl. no	Name	DNA offtarget target seq with PAM	Chromo some	Position	Gene location	Mis matches
1	Off-T-1	GATaTGGTAA gAGTCTAGGa	chrY	22321103	no track data	3
2	Off-T-2	GATCTGGaA gCAGTtTAGGC	chr11	116360344	no track data	3
3	Off-T-3	GATtTGGTAA CAGTCTctGC	chr18	14965946	Intron of LINC01443	3
4	On-T	GATCTGGTAA CAGTCTAGGC	chr17	44386498	Intron of ITGA2B	3

Table 3.2.10: Primers to amplify DNA off-target

Sl. no	Primer name	Primer sequence 5'- 3'	Amplicon size	Source
1	IT6 OT1 F	tacacgacgctcttccgatct aagtggaagttcttacagagatgcatttcagg	215bp	This study
2	IT6 OT1 R	agacgtgtgctcttccgatct gcacgtgaactgaagttctctcc		This study
3	IT6 OT2 F	tacacgacgctcttccgatct ggcaggggaataacttggg	200bp	This study
4	IT6 OT2 R	agacgtgtgctcttccgatct taacaccctcggtggctcct		This study
5	IT6 OT3 F	tacacgacgctcttccgatct acttgctggagatcccatggtg	200bp	This study
6	IT6 OT3 R	agacgtgtgctcttccgatct gcaagtagagacacagatgggtca		This study

3.2.15 Statistics:

Statistical analyses in this study were performed using GraphPad Prism 8.1. Unpaired two-sided t-tests were employed for independent group comparisons, assuming a normal distribution. Multiple group comparisons utilized one-way analysis of variance (ANOVA). For repeated measures within groups, ordinary or repeated measure ANOVA was applied based on the experimental design. Correlation analysis was used to explore sample relationships, and Principal Component Analysis (PCA) was conducted to identify underlying patterns. A significance level of $p < 0.05$ was considered for all statistical tests conducted.

RESULTS

Section 4.1: Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin.

Previous research has demonstrated the crucial role of the highly similar HBG1 and HBG2 proximal promoters in the expression of gamma-globin. The occurrence of certain point mutations in the promoter region of these genes has been associated with non-deletional forms of hereditary persistence of fetal hemoglobin (HPFH), which has been linked to an increase in gamma-globin expression (Wienert *et al*., 2015). To identify novel regulatory elements that modulate gamma-globin expression in these promoters, a base editing screen was conducted by introducing point mutations in all appropriate locations within 320 base pairs upstream of the transcription start site of the HBG genes.

To accomplish this, a stable HUDEP-2 cell line, which is an immortalized human erythroid progenitor cell line, was established (Kurita *et al*., 2013). These cells expressed base editors (either ABE or CBE). Subsequently, guide RNAs (gRNAs) that targeted the proximal promoter region of HBG1 and HBG2 were screened to assess their ability to upregulate fetal globin expression. The most effective gRNAs were evaluated for their editing efficiency and HbF levels, and the potential mechanism of the newly identified gRNAs in HbF elevation was further investigated using EMSA and ChIP qPCR. Finally, the capacity of the newly identified gRNAs to induce fetal hemoglobin levels was confirmed in erythroid cells derived from healthy donor CD34+ hematopoietic stem and progenitor cells (HSPCs), thus demonstrating their potential as a therapeutic approach.

4.1.1 Base editors as a preferred genome editing tool for targeting the highly homologous *HBG* promoter region.

Initially, we generated stable HUDEP-2 cell lines that express various gene editors, such as ABE, CBE, or Cas9. HUDEP-2 cells were transduced with ABE7.10 RA, BE3RA-FNLS, or Cas9 lentiviral constructs, hereafter referred to as HUDEP-2-ABE, CBE, or Cas9. The vector copy number (VCN) of HUDEP-2 cells transduced with ABE, CBE, or Cas9 lentiviral constructs ranged from 0.25 to 0.85, as determined by real-time PCR (**Fig 4.1.2a**). Previously defined sgRNAs that targeted the BCL11a binding motif in the HBG1 and HBG2 promoters with a suitable editing window for ABE, CBE, and Cas9 were transduced with a VCN of 0.6 to 1.2 (**Fig 4.1.2a**). The editing efficiency was 88% for Cas9, 10-51% for ABE, and 59-73% for CBE, with the transduction efficiency (as measured by GFP expression) exceeding 98% (**Fig 4.1.1a**). After differentiation of the cells into erythroid progenitors, the percentage of HbF-positive cells was greater for ABE and CBE than for Cas9 (**Fig 4.1.1b**). Despite high levels of gene editing with Cas9, we did not observe a corresponding increase in HbF-positive cells, which may be due to a previously described 4.9 kb deletion comprising the HBG2 gene and HBG1-HBG2 intergenic region. Introducing double-strand breaks (DSBs) in highly homologous HBG promoters with Cas9 nucleases can result in this deletion (Li *et al*., 2018). Therefore, we determined the frequency of this deletion by qRT-PCR in all edited samples. As expected, the 4.9 kb deletion was observed at a high frequency (76%) in Cas9-edited cells. We also detected some deletions in base-edited samples, but significantly fewer than with Cas9 (**Fig 4.1.1c**). Consistent with the frequency of the 4.9 kb deletion, globin chain analysis by RP-HPLC revealed lower levels of G gamma, but not A gamma chain, only in Cas9-edited cells compared to ABE and CBE-edited cells after normalization to the control (**Fig 4.1.1d**). These findings

suggest that ABE and CBE are highly efficient in editing highly homologous regions such as the gamma-globin promoter without causing a large deletion between the two HBG genes.

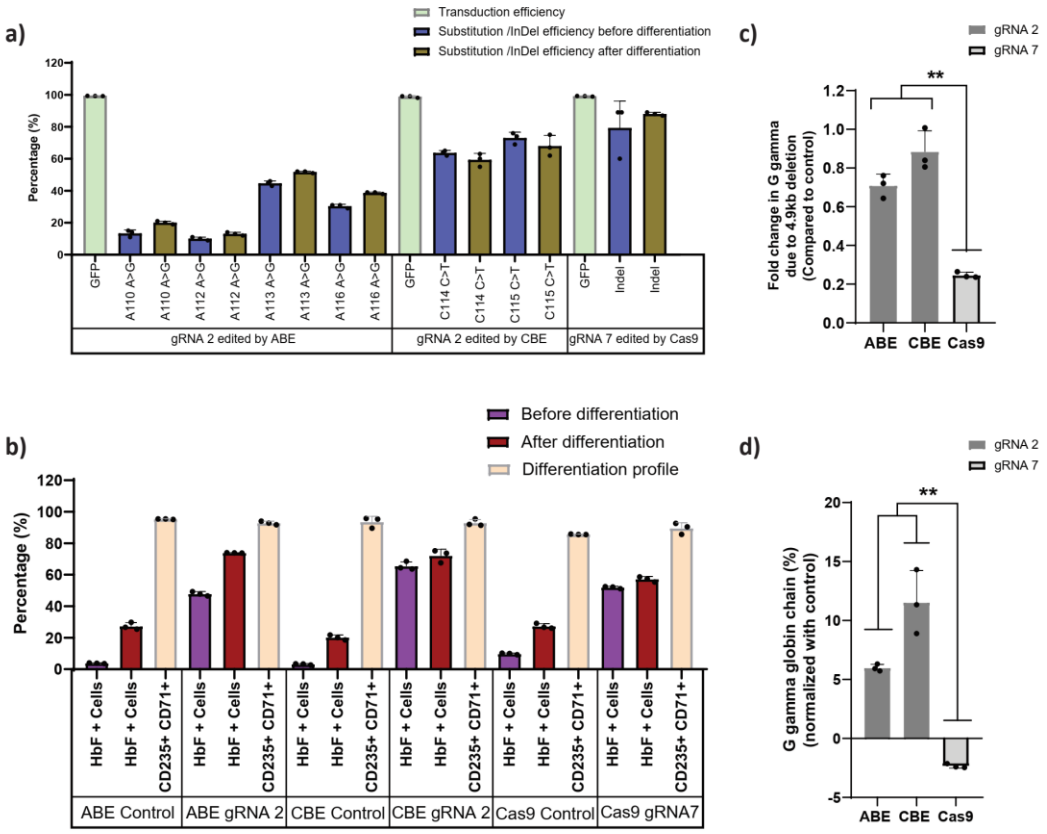


Figure 4.1.1: Base editors were chosen over Cas9 for editing the highly homologous HBG1 and HBG2 promoter. We employed adenine base editor (ABE), cytosine base editor (CBE), and Cas9 with suitable guide RNAs (gRNAs) that target the well-known BCL11A binding site (-115 transcription start site [TSS]). (a) The transduction efficiency of gRNA-2 (for ABE and CBE) or gRNA-7 (for Cas9), percentage of individual base conversion for ABE and CBE (with gRNA-2) and insertions and deletions (indels) for Cas9 (with gRNA-7) were analyzed before and after erythroid differentiation. The transduction efficiency was analyzed by fluorescence-activated cell sorting (FACS), and the individual base substitution and indel percentage were analyzed by EditR and ICE software respectively after Sanger sequencing. (b) Flow cytometry analysis of fetal hemoglobin (HbF) and erythroid maturation markers (CD235a and CD71) expression in edited HUDEP-2 cells were performed. The percentage of HbF-expressing cells were analyzed before and after differentiation into erythroblasts. (c) Analysis of HBG2 deletion (due to 4.9 kb deletion) was carried out by quantitative reverse transcription PCR (qRT-PCR) in the base-edited and Cas9-edited HUDEP-2 cells. (d) The expression of G gamma-globin chain in ABE, CBE, and Cas9-edited HUDEP-2 cells was measured by reverse-phase high-performance liquid chromatography (RP-HPLC) after differentiation into erythroblasts. The data were normalized with respective controls. The data are expressed as mean ± standard error of the mean (SEM) from three biological replicates, and statistical significance was indicated by asterisks (**$p < 0.01$).

4.1.2 Screening of *HBG* proximal promoter with base editors identifies novel HPFH like mutations.

To investigate the potential regulatory regions associated with gamma-globin expression, we utilized ABE and CBE for tiling the HBG promoter. Through transcriptomic analysis of stable cell lines expressing ABE and CBE, we confirmed that the gene expression profiles were not altered and showed a significant correlation with wild-type HUDEP-2 cells (**Fig 4.1.2b**). As the base editors and gRNAs are constitutively expressed, we assessed the editing frequency of ABE and CBE stables with gRNA-2 at different time points during expansion and differentiation to determine its effect on HbF elevation. We observed an increase in editing efficiency and HbF levels in both ABE and CBE over time, with no discernable impact on erythroid differentiation (**Fig 4.1.2c-h**).

We then generated ABE and CBE gRNAs in all compatible locations up to 320 bp upstream of the transcription start site (TSS) of the HBG1 and HBG2 promoters. The guide RNAs were designed with a suitable base editing window, targeting nucleotides in position 3-9 from the NGG PAM distal end for ABE and CBE (**Fig 4.1.3a**). Out of the 41 gRNAs designed, 36 gRNAs had a base editing window for ABE, and 32 gRNAs had a base editing window for CBE. An overview of the methodology used in this study is illustrated in **Figure 4.1.3b**. All the gRNAs were cloned in a lentiviral vector with a GFP reporter, and lentivirus was produced for each gRNA. HUDEP-2-ABE and -CBE cells were then transduced in an arrayed format with equal transduction efficiency (~1 VCN/cell). The mean transduction efficiency for all these gRNAs in both ABE and CBE samples was around 97%. The gRNA transduced cells were then expanded for eight days, and successful base editing was confirmed by NGS and Sanger sequencing.

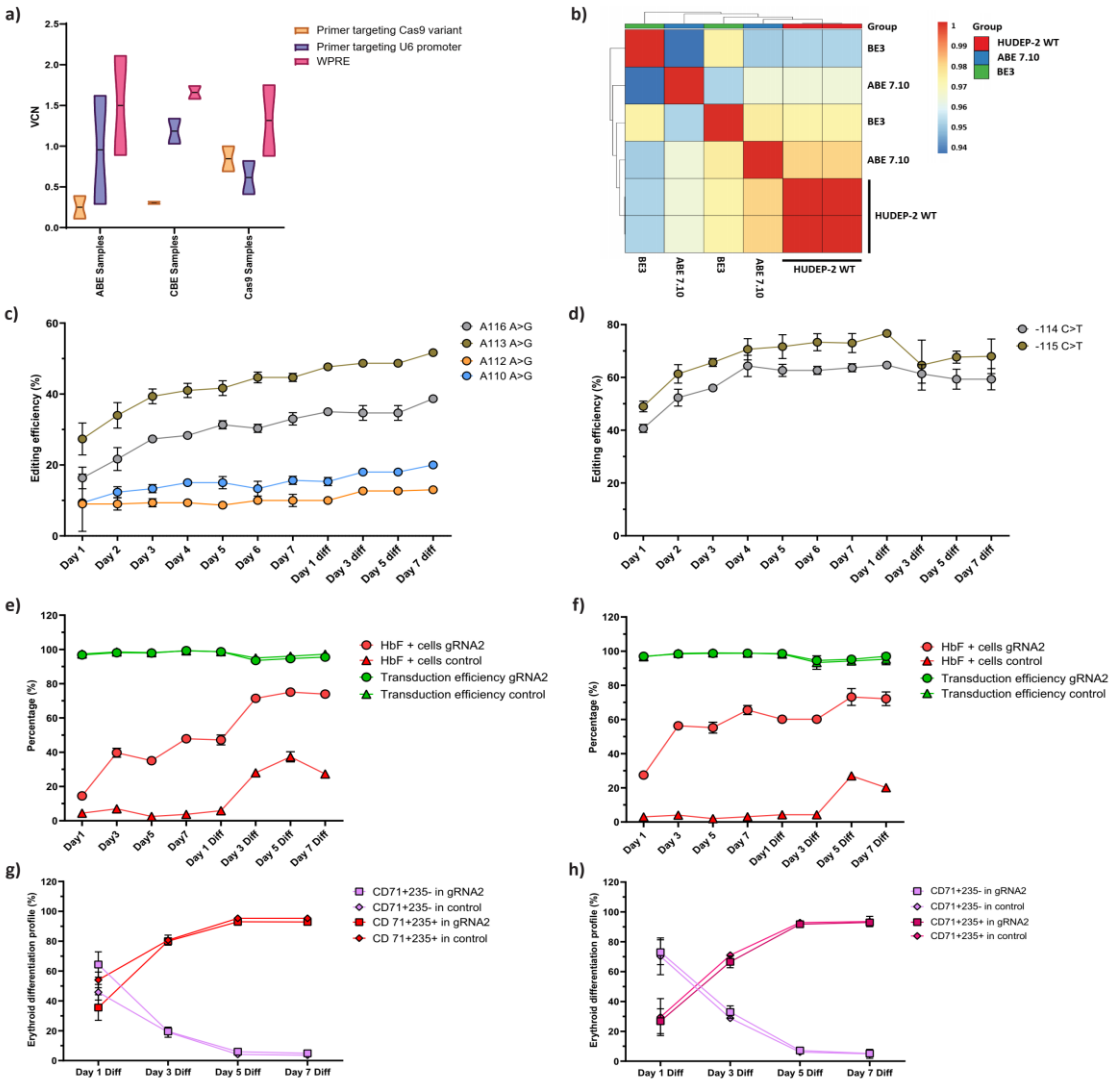


Figure 4.1.2: The HUDEP-2 cells expressing adenine base editor (ABE) and cytosine base editor (CBE) were characterized as follows: (a) The vector copy number (VCN) per cell for the integrated guide RNA (gRNA) and the gene editors (ABE, CBE, and Cas9) were determined in the respective HUDEP-2 stable cell lines. Primers specific to the Cas9 gene editors and the U6 promoter for gRNA were used, while the primer targeting WPRE was common for both gRNA and gene editors. (b) Transcriptome analysis of ABE and CBE expressing HUDEP-2 cells was conducted, and the results were represented in a pairwise Pearson correlation matrix. The R value for individual boxes was represented by the gradient bar. (c) The individual base conversion efficiency of ABE or CBE expressing HUDEP-2 cells transduced with gRNA-2 on different days of expansion and differentiation was determined. The base substitution at the BCL11A binding site of the HBG promoter (-115 region) was analyzed by EditR after Sanger sequencing. (d) Flow cytometry analysis of fetal hemoglobin (HbF) and GFP expression in ABE or CBE expressing HUDEP-2 cells transduced with gRNA-2 during different days of expansion and erythroid differentiation was performed. The expression of GFP was directly proportional to the percentage of cells transduced with gRNA-2. (e) The differentiation profile of the base edited HUDEP-2 cells expressing ABE and CBE was determined from day 1 to day 7 of erythroid differentiation, measured by flow cytometry. The decrease in the CD71+ CD235- population and increase in the CD71+ CD235+ population displayed progress in erythroid differentiation. The data are presented as mean ± SEM from three biological replicates.

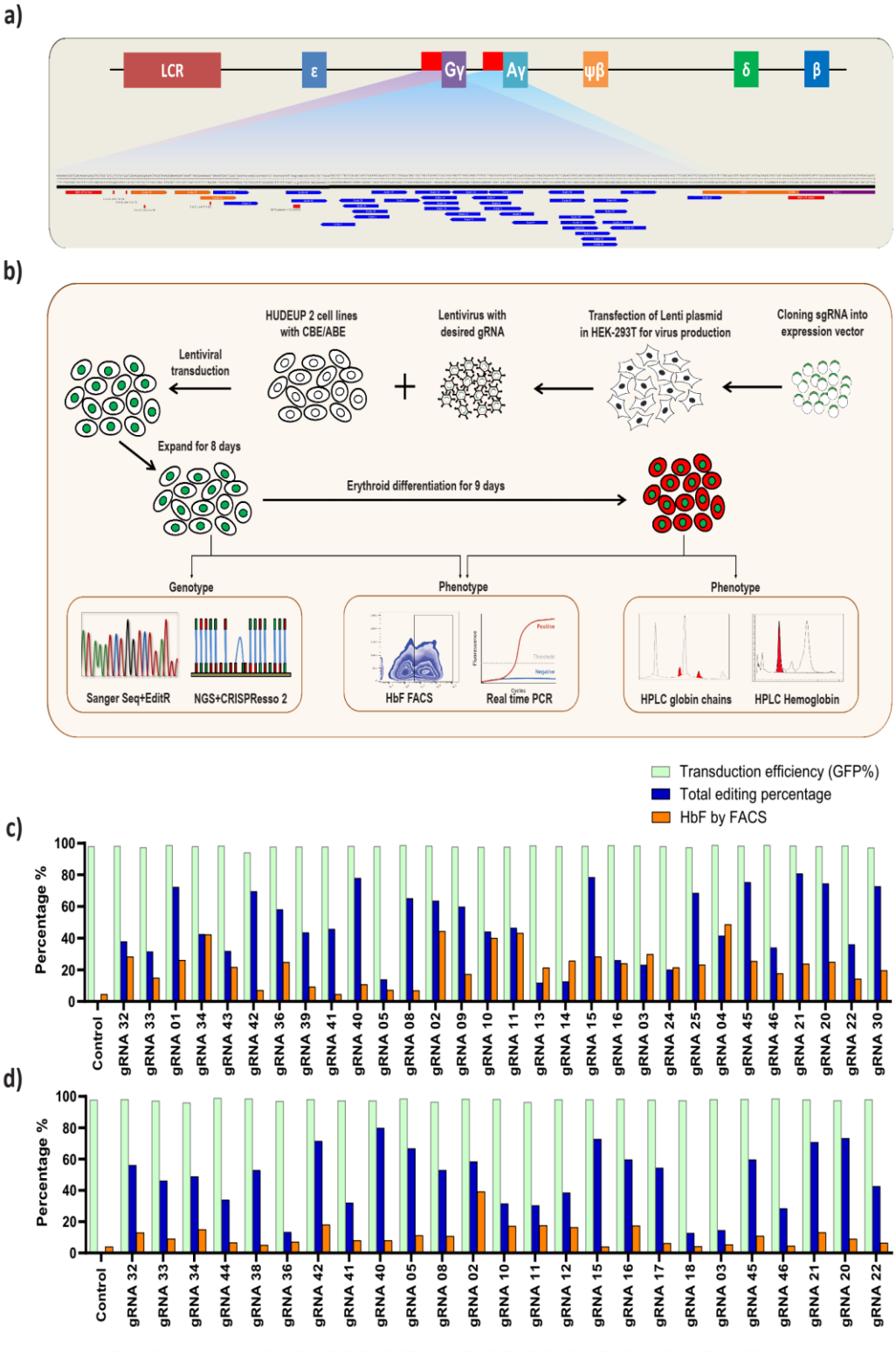


Figure 4.1.3: Identification of novel point mutations that enhance fetal hemoglobin (HbF) expression via screening of the HBG promoter using base editors. (a) This figure illustrates the screening methodology used in this study. HUDEP-2 cells, which express either an adenine base editor (ABE) or a cytosine base editor (CBE), were genetically modified with guide RNAs (gRNAs) that specifically target the proximal promoter region of the HBG gene. Following editing, the cells were expanded for eight days, and the editing efficiency was assessed using both Sanger sequencing and Next-

Generation Sequencing (NGS). To assess functionality, FACS and quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) were employed. The most promising targets from both ABE and CBE screens, with the highest HbF induction, were further validated and differentiated into erythroid cells. FACS, qRT-PCR, Reverse-Phase High-Performance Liquid Chromatography (RP-HPLC), and High-Performance Liquid Chromatography (HPLC) analyses were carried out on these differentiated cells to evaluate HbF expression levels, HBG gene expression, individual gamma-globin chains, and fetal hemoglobin levels. (b) Depiction of guide RNAs (gRNAs) targeting the HBG promoter region in the HUDEP-2 cell line. The gRNAs were designed to target the region -320 bp upstream of the transcription start site (TSS) in both HBG1 and HBG2 promoter regions, and are displayed in the figure. The gRNAs common to both HBG1 and HBG2 promoters are shown in blue, while the gRNAs specific to the HBG1 promoter are in orange. The red bar represents the primers used for deep sequencing. The transduction efficiency, base editing frequency, and HbF expression in HUDEP-2 cells expressing adenine base editor (ABE) (c) and cytosine base editor (CBE) (d) that were transduced with various gRNAs targeting the HBG proximal promoter were compared. The edited cells were subjected to next-generation sequencing (NGS) to determine the total editing frequency using CRISPResso-2. The transduction efficiency (GFP+ cells) and the number of HbF positive cells were assessed by fluorescence-activated cell sorting (FACS).

In this study, we utilized Next-Generation Sequencing (NGS) to evaluate the efficacy of ABE-induced A to G conversions and CBE-induced C to T conversions at various target sites of the HBG promoter. We excluded gRNAs associated with low base editing efficiency (<10%) from further analysis, as they do not provide meaningful insights into HBG regulation (gRNAs -37, 38, 7, 18, 19, 29 in ABE and gRNAs -1, 35, 37, 6, 7, 13, 19 in CBE) (data not shown). The total base editing efficiency of the remaining gRNAs varied from 12-81% and 13-80% for ABE (n=30) and CBE (n=25), respectively, as determined by CRISPResso-2 analysis (**Fig 4.1.3 c-d**). We also examined the individual base conversion frequency of ABE and CBE, which ranged from 0-74% and 0-61%, respectively (**Fig 4.1.4 a & c**). Sanger sequencing data analyzed by EditR further confirmed the base substitution efficiency at the target loci (**Fig 4.1.4 b & d**). The base substitution efficiency varied significantly depending on the base editing window for ABE and CBE, with high average editing efficiency observed in the canonical positions (A5-A7 for ABE and C5-C7 for CBE) and lower efficiency in the non-canonical positions (A1-A4, A8-A12 for ABE and C1-C4, C8-C16 for CBE) for the different gRNAs used in this study (**Fig 4.1.5 e-f**).

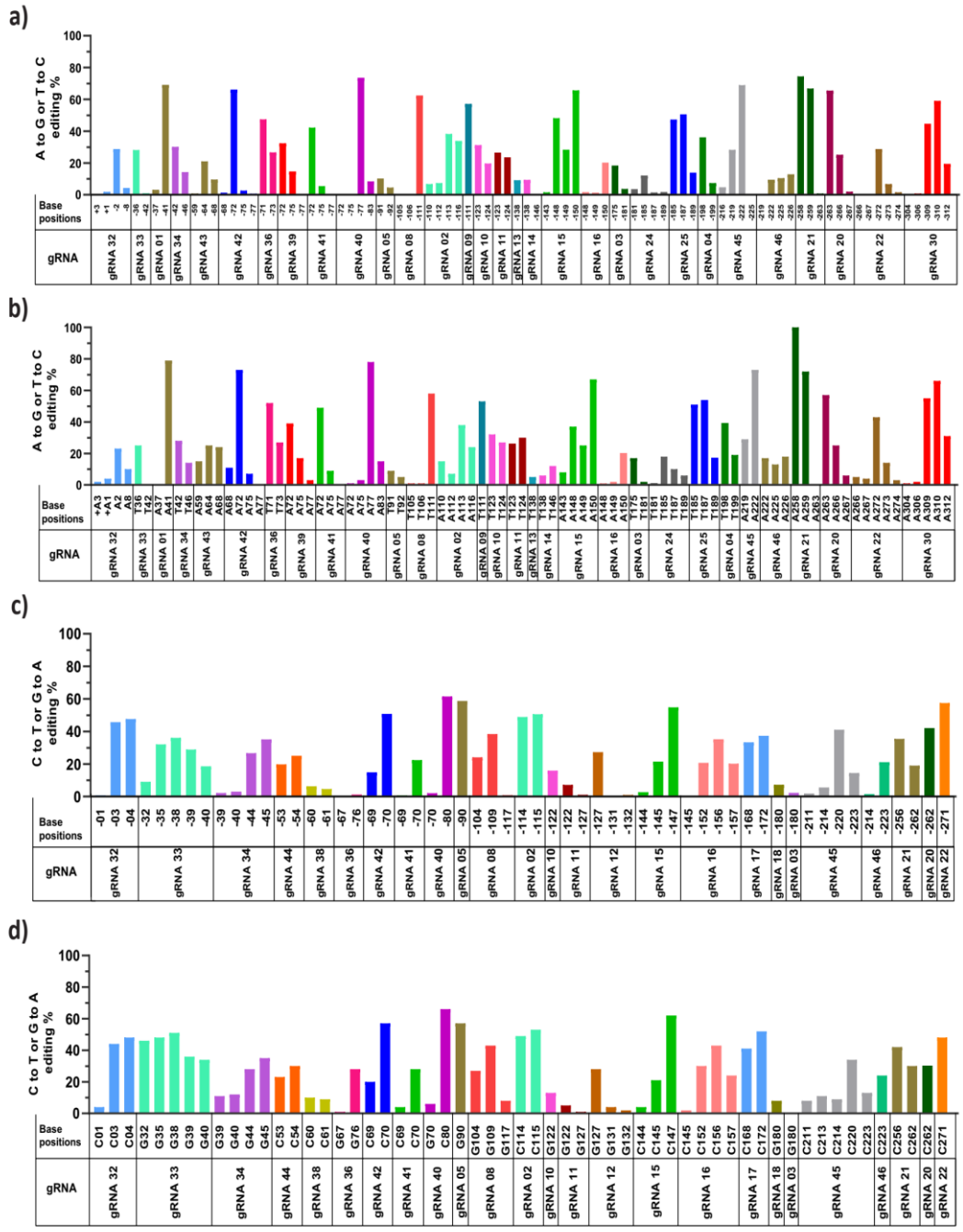


Figure 4.1.4: The efficacy of adenine base editor (ABE) and cytosine base editor (CBE) in inducing base substitutions in the HBG promoter was assessed at the single base pair level using next-generation sequencing (NGS) and Sanger sequencing. Stable cells expressing ABE and CBE and transduced with specific guide RNAs (gRNAs) were sequenced using NGS and Sanger sequencing, respectively (figure (a) ABE and (c) CBE for NGS, and figure (b) ABE and (d) CBE for Sanger sequencing). Sequencing data were analyzed using CRISPResso-2 and EditR software packages, respectively. The positions of base conversion were sequentially arranged in the x-axis up to –320 bp from the transcription start site (TSS), and the base conversion efficiency for each gRNA was plotted in the y-axis as the percentage of the desired base conversion (A-to-G or C-to-T).

subsequently, we investigated the product purity and indel percentage for all gRNAs in ABE and CBE, as previous research has indicated that base editors generate unintended edits at target sites with low frequency (Koblan *et al*., 2018). Our findings demonstrate that most of the gRNAs in CBE transduced cells produced unexpected C-to-non-T edits (C-R/G-Y), with C-to-G conversion being the most prevalent. In contrast, we observed minimal levels of unexpected base conversions (A-Y/T-R) in a few on-target sites in ABE, consistent with previous studies (Nicole M Gaudelli *et al*., 2017; Komor, Kim, Packer, Zuris and Liu, 2016)(**Fig 4.1.5 a -b**). The indel frequency was less than 2% in both ABE and CBE, as determined by deep sequencing data (**Fig 4.1.5 c-d**). Our results indicate that ABE exhibits higher product purity and lower indel frequency than CBE in all cases. Overall, our study demonstrates that both ABE and CBE can effectively introduce A-to-G and C-to-T nucleotide substitutions, respectively, in the proximal promoters of HBG1 and HBG2.

To assess the impact of ABE and CBE-mediated targeted base substitution at the HBG promoter on HbF expression, we employed flow-cytometry to analyze HbF-positive cells following intracellular HbF staining. The observed percentage of HbF-positive cells ranged from 2-44% in ABE and 1-35% in CBE (**Fig 4.1.3c-d**). Among the 30 ABE and 25 CBE gRNAs screened for editing the HBG promoter, five ABE gRNAs and one CBE gRNA exhibited a notable increase in the number of HbF-positive cells (ranging from 40-50%). Our subsequent investigations revealed that some of the top gRNA candidates included previously identified binding sites for BCL11A (gRNA-2), KLF-1 (gRNA-4), and TAL-1 (gRNA-3), along with several novel target sites (gRNA-10, gRNA-11, gRNA-15, gRNA-16, gRNA-21, gRNA-32, gRNA-34, gRNA-42).

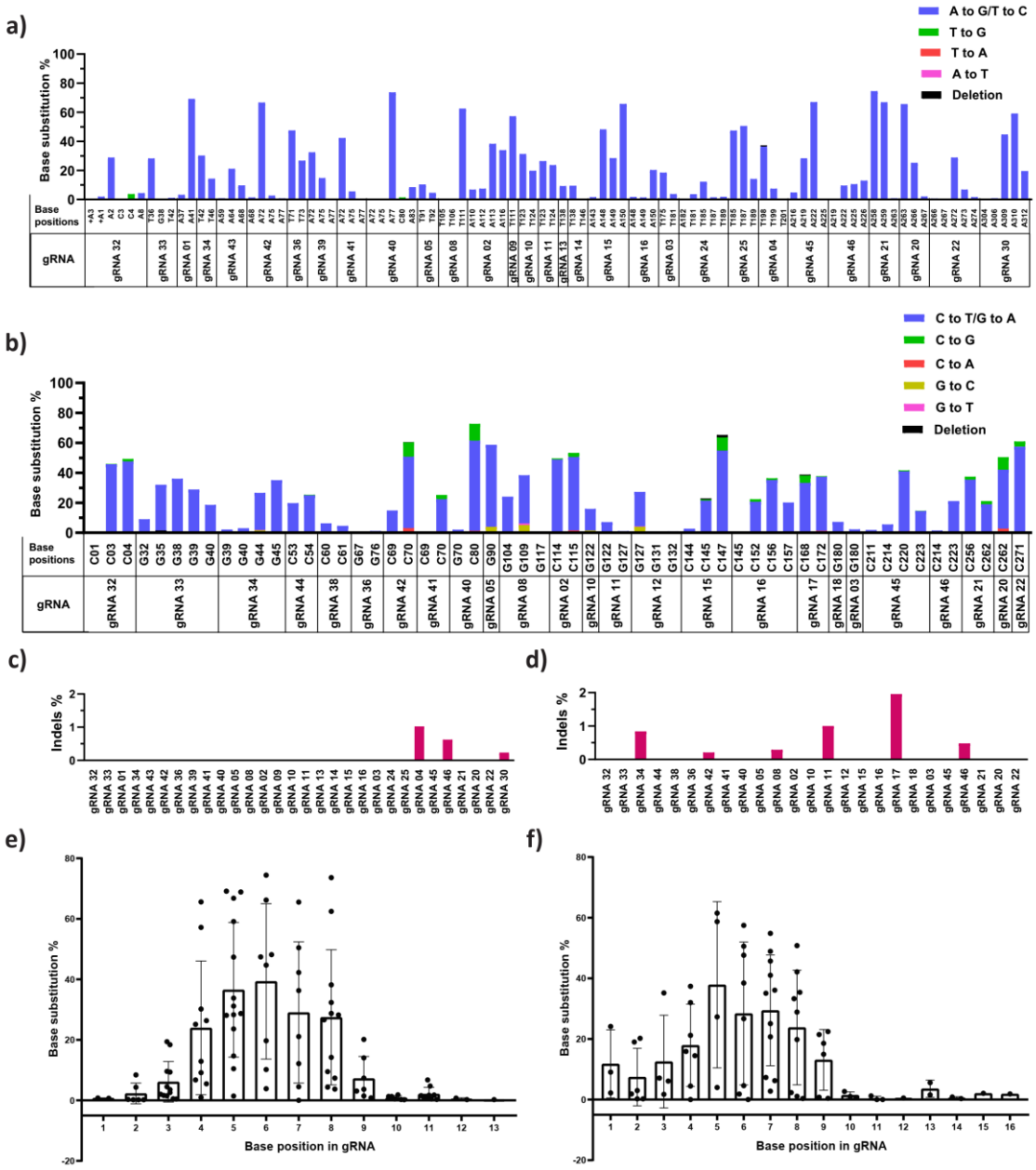


Figure 4.1.5: Analysis of Product Purity and Optimal Editing Window of Adenine (ABE) and Cytosine Base Editors (CBE) at Target Sites. The specific and non-specific editing patterns of adenine base editor (ABE) (a) and cytosine base editor (CBE) (b) were assessed at target sites with different guide RNAs (gRNAs) using deep sequencing (NGS). The individual base positions within the HBG promoter region were plotted on the x-axis, while the y-axis represented the number of reads showing specific and non-specific conversion as a percentage. The CRISPResso-2 tool was utilized to analyze the percentage of insertions and deletions (indels) at specific target sites in cells treated with ABE (c) and CBE (d) after NGS. The effect of the position of a target base on the editing scope of ABE (e) and CBE (f) for different gRNAs was also investigated. The x-axis represents the base position in the protospacer, with the PAM considered as positions 21–23, while the y-axis represents the percentage of base substitution.

In addition, we explored whether the gRNAs with higher total editing efficiency (>40%) but lower HbF levels (<10%) were affecting activator binding sites, thereby down-regulating gamma-globin expression. Using K562 cell lines stably expressing ABE or CBE, we transduced selected gRNAs and achieved higher individual base editing efficiency for each gRNA (**Fig 4.1.6 a-b**). However, we did not observe any decrease in the number of HbF-positive cells in any of the samples even with high transduction rate, indicating that the targeted regions did not contain binding sites for essential transcriptional activators (**Fig 4.1.6 c-d**).

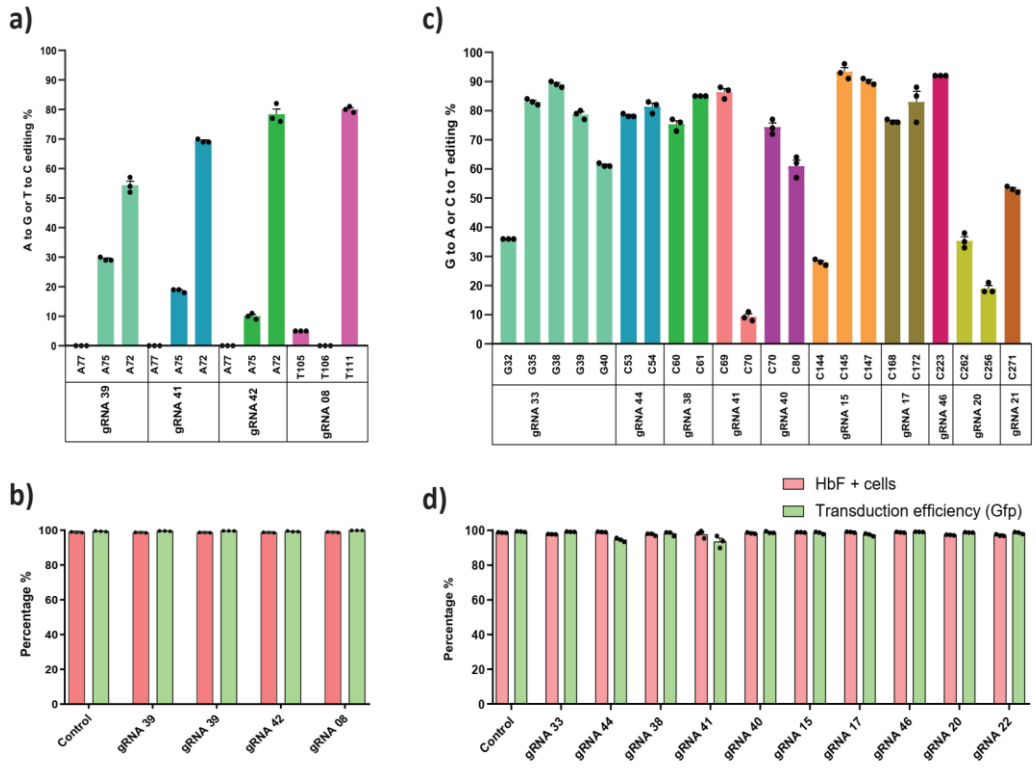


Figure 4.1.6: Identification of Nucleotide Substitutions Suppressing Fetal Hemoglobin (HbF) Expression through Base Editing of the HBG Promoter. K562 cells expressing adenine base editor (ABE) (a) or cytosine base editor (CBE) (b) were transduced with the corresponding guide RNA (gRNA), and the resulting sequences were obtained using Sanger sequencing and analyzed with EditR. Based on the initial screening results, the gRNAs that showed lower HbF levels with higher editing efficiency were chosen. The percentage of HbF-positive cells and the transduction efficiency (GFP+) of each gRNA for ABE (c) and CBE (d) were quantified using flow cytometry.

Notably, several of the top contenders from the screening process included regions of interest that were previously identified as binding sites for BCL11A (gRNA-

2), KLF-1 (gRNA-4), and TAL-1 (gRNA-3). In addition, we also discovered several other novel target sites (gRNA-10, gRNA-11, gRNA-15, gRNA-16, gRNA-21, gRNA-32, gRNA-34, gRNA-42). Of note, gRNAs-2, -3, and -4 mimic well-known naturally occurring HPFH mutations –114C > T, –117G > A, −175T > C, and –198T > C (Liu *et al*., 2018)(Martyn *et al*., 2018)(Stoming TA *et al*, 1989)(Wienert *et al*., 2017, 2018). We compared the HbF positive cell percentage with the editing efficiency of each gRNA at the target site in both ABE and CBE cells (**Fig 4.1.3c–d**). The overall base editing efficiency was typically higher than the proportion of HbF positive cells, with the exception of a few cases (gRNA-3, -4, -13, -14, -37, and -38 in ABE edited cells). Overall, the candidate gRNAs that increased HbF expression identified from the primary screening of the HBG promoter by ABE and CBE provide targets for further validation.

4.1.3 Enhanced HbF Expression through Base Editing at Potential Target Sites in the HBG Promoter:

The top eight gRNAs (2, 3, 4, 10, 11, 15, 32, and 34) identified from the ABE screen and the CBE screen (2, 10, 11, 16, 21, 32, 34, and 42), which yielded the highest levels of HbF-positive cells, underwent further validation. Among these gRNAs, five (2, 10, 11, 32, and 34) were common to both ABE and CBE, suggesting that these target regions likely play a crucial role in the silencing of the HBG gene. Following initial expansion, the edited cells were cultured in erythroid differentiation media, and a battery of functional assays was conducted (**Fig 4.1.3b**).

Consistent with the screening results, the overall editing efficiency ranged from 24% to 78% for ABE and from 36% to 85% for CBE, with mean transduction efficiencies of 96% and 90%, respectively (**Fig 4.1.7a-b and Fig 4.1.8f**). At the target

regions, we observed individual base conversions from A to G (ranging from 0% to 65%) or from C to T (ranging from 1% to 57%) with indel frequencies below 2% (**Fig 4.1.8a-b and e**). Furthermore, undesired non-C-to-T conversions (i.e., C to A or C to G) were observed at the on-target site with CBE but not with ABE (**Fig 4.1.8a-b and Fig 4.1.9a-b**). The specific nucleotide substitution patterns mediated by ABE or CBE for all eight top gRNAs are illustrated in the corresponding figures (**Fig 4.1.9a-b**). ABE exhibited higher efficiencies in editing the cognate bases (A and T) (A113 and A116 for gRNA-02, T175 for gRNA-03, T198 for gRNA-04) compared to the bystander bases (A110 and A112 for gRNA-02, T181 for gRNA-03, T199 for gRNA-04) for generating HPFH mutations. In the case of CBE, we also observed base conversion from C to T in nucleotides adjacent to the protospacer sequence, as previously reported (Arbab *et al*., 2020; Webber *et al*., 2019). For instance, gRNA-10 and gRNA-11 in CBE exhibited base conversion outside the protospacer sequence (-117 site) in addition to on-target editing at the -122 site within the base editing window (**Fig 4.1.8b and Fig 4.1.9b**). The base conversion at the -117 site disrupted the core binding motif of the major fetal globin repressor-BCL11A (Martyn *et al*., 2018; Wienert *et al*., 2018). To distinguish the editing frequency in HBG1 and HBG2 promoters, we phased the edits with single nucleotide variations at positions -271, -307, -317, and -324, which are unique to each promoter. This analysis, utilizing Bowtie 2 and IGV software, revealed highly similar base editing rates and no variation in base substitution efficiency between the highly homologous HBG1 and HBG2 promoters (James T. Robinson *et al*, 2012) (**Fig 4.1.8c-d**).

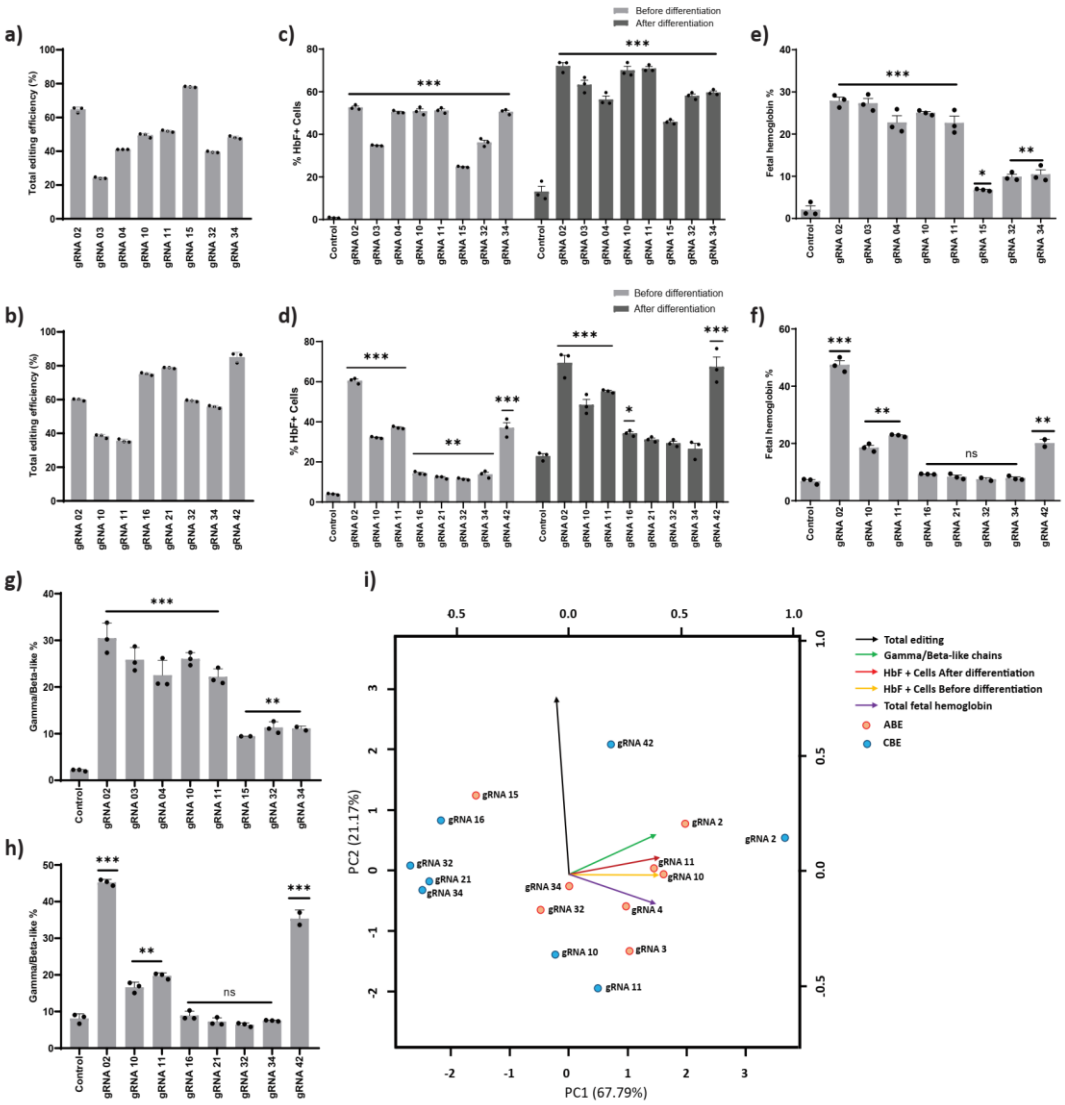


Figure 4.1.7: Verification of Targeted Base Editing Efficiency for the Most Promising Eight Guide RNAs (gRNAs) Identified in the Initial Screening of Adenine Base Editor (ABE) and Cytosine Base Editor (CBE) on HBG Promoters. HUDEP-2 cells expressing Adenine Base Editor (ABE) or Cytosine Base Editor (CBE), transduced with the top eight guide RNAs (gRNAs), underwent deep sequencing analysis targeting specific regions within the HBG promoter. The editing efficiencies of ABE (a) or CBE (b) were quantified as the percentage of total sequencing reads exhibiting the desired conversion of C:A to T:G at predetermined sites. The assessment of fetal hemoglobin (HbF) positive cells in HUDEP-2 cells expressing ABE (c) or CBE (d), transduced with the respective gRNAs, was performed before and after differentiation using flow cytometry. Analysis of globin chains in ABE (e) or CBE (f) edited HUDEP-2 cells was carried out following erythroid differentiation using RP-HPLC. Additionally, HbF analysis in ABE (g) or CBE (h) edited HUDEP-2 cells was conducted post-erythroid differentiation using HPLC. A principal component analysis (PCA) plot (i) was generated to investigate the correlation between the outcomes of base editing using the top eight gRNAs. The relationship between base edit frequency, HbF+ cells, HbF levels, and gamma/beta-like chains in ABE or CBE edited HUDEP-2 stable cells was analyzed for the indicated gRNAs. The plot displays the first two principal components, illustrating the variance accounted for by each component. The presented data is expressed as mean ± SEM from three biological replicates ($p > 0.05$). Asterisks indicate levels of statistical significance, with **$p < 0.01$ and ***$p < 0.001$.

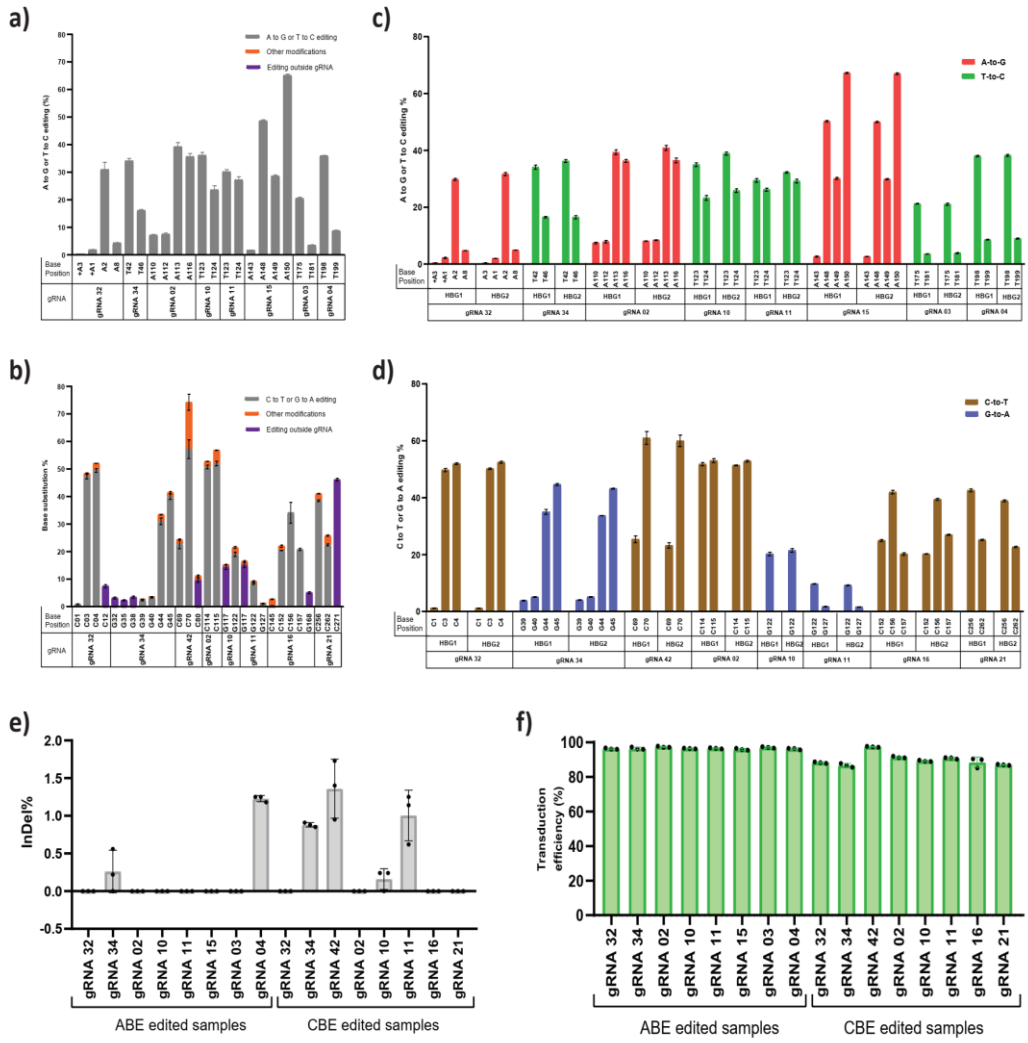


Figure 4.1.8: Evaluation of Base Editing Efficacy at the Highly Homologous Promoters of HBG1 and HBG2. HUDEP-2 cells expressing Adenine Base Editor (ABE) or Cytosine Base Editor (CBE) were transduced with the top eight guide RNAs (gRNAs) and subjected to deep sequencing analysis targeting specific regions within the HBG promoter. The base substitutions introduced by ABE (a) and CBE (b) at single base resolution were depicted for each respective gRNA at the target site. Unintended conversions at the target site and conversions occurring beyond the protospacer were represented by orange and blue bars, respectively. The base editing efficiencies of ABE (c) and CBE (d) were compared at the designated target sites within the HBG1 and HBG2 promoter regions in HUDEP-2 cells. To distinguish specific editing events between the highly homologous HBG1 and HBG2 promoters, amplification and deep sequencing of the combined HBG1 and HBG2 promoter regions were performed, utilizing four single nucleotide variations at positions −271, −307, −317, and −324. Furthermore, the frequency of indels (insertions or deletions) (e) and the transduction efficiency (f) were represented for ABE and CBE stable cells transduced with the top eight gRNAs. The presented data is expressed as mean ± SEM from three biological replicates.

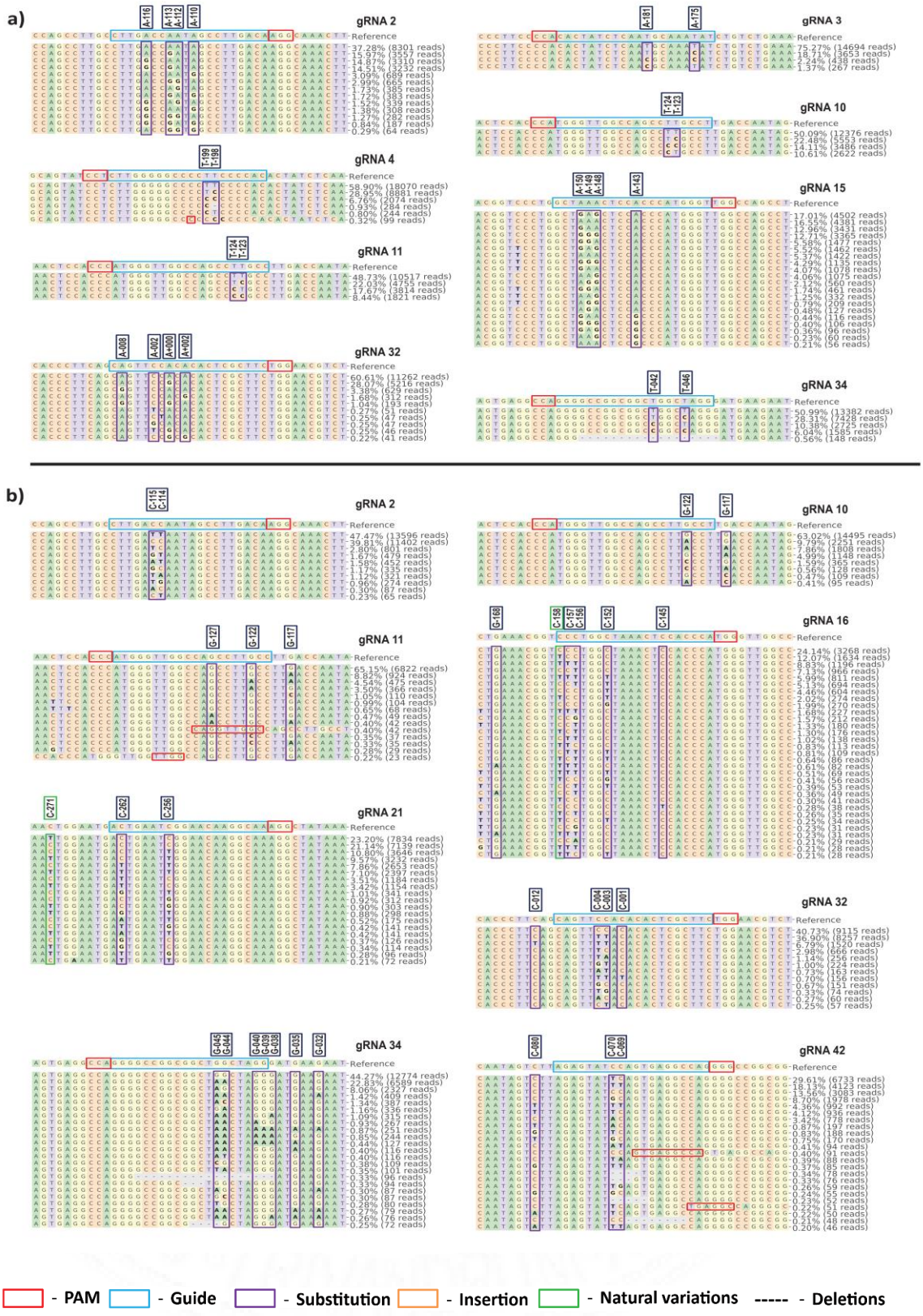


Figure 4.1.9: Overview of Allele Frequencies at the Targeted Sites for the Top Eight Guide RNAs (gRNAs) by Adenine Base Editor (ABE) and Cytosine Base Editor (CBE). The targeted region within the proximal promoter of HBG (Human β-globin) was amplified via PCR and subsequently subjected to next-generation amplicon sequencing for analysis. The overall frequencies of base editing, including

substitutions, insertions, natural variations, and deletions, were determined for each gRNA in ABE (a) and CBE (b) edited cells. The gRNA sequence, PAM (Protospacer Adjacent Motif) site, specific base substitutions, and their positions within the HBG promoter of HUDEP-2 cells are provided in the corresponding boxes for reference.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis was conducted to examine the expression of HBG (fetal hemoglobin) before and during the differentiation process. A significant increase in HBG mRNA expression was observed for all top eight gRNAs in ABE-edited cells ($p<0.01$-$p<0.0001$), indicating a robust induction of HBG expression (**Fig 4.1.10 a**). In the case of CBE, gRNAs 2, 10, 11, and 42 exhibited a substantial increase in HBG mRNA expression ($p<0.05$-$p<0.0001$), while gRNAs 16, 21, 32, and 34 showed a modest level of expression compared to the control (**Fig 4.1.10g**) before differentiation. The expression pattern of globin mRNA during erythroid differentiation showed a similar trend for both ABE (**Fig 4.1.10b**) and CBE (**Fig 4.1.10h**) edited cells.

To evaluate the impact of base editing on HbF production, the percentage of HbF-positive cells was determined before and after erythroid differentiation using fluorescence-activated cell sorting (FACS). As anticipated, the differentiated erythroid cells exhibited a slightly higher percentage of HbF-positive cells compared to the undifferentiated edited cells (**Fig 4.1.7c-d**). Furthermore, the effect of base editing on erythroid differentiation was assessed by flow cytometry analysis using CD235a and CD71 markers. The transition from CD71-positive cells alone to CD71/CD235a double-positive cells indicates the erythroid differentiation pattern of HUDEP-2 cells (Kurita *et al*., 2013). The percentage of double-positive cells was 83-90% for CBE edited cells and above 95% for ABE edited cells, whereas the control cells exhibited 77% and 97% double-positive cells, respectively. These findings suggest that the differentiation capacity of the edited cells was not significantly affected (**Fig 4.1.10d and Fig 4.1.10j**).

Moreover, the abundance of globin chains was assessed using reverse-phase high-performance liquid chromatography (HPLC) in differentiated erythroid cells derived from both ABE and CBE edited samples. A notable increase in gamma-globin chain expression was observed, constituting 10-30% of the total beta-like globin content in all ABE edited samples (gRNAs 2, 3, 4, 10, 11, 15, 32, and 34) (**Fig 4.1.7g**). In the case of CBE, the gamma-globin chain levels ranged from 6% to 45% of the total beta-like globin content, with gRNAs 2, 42, 10, and 11 exhibiting significant elevation compared to the control (**Fig 4.1.7h**). Concurrently, a reciprocal reduction in beta-globin chains was observed, maintaining the alpha to beta-like globin chain ratio in both ABE and CBE edited cells (**Fig 4.1.10c and Fig 4.1.10i**). Despite comparable editing efficiency in the HBG1 and HBG2 promoters, HBG1 demonstrated moderately higher expression levels compared to HBG2 in most of the ABE and CBE edited cells. The decreased expression of HBG2 in these samples could be attributed to biased HbF regulation or the presence of the 4.9kb deletion that removes the HBG2 gene.

To investigate whether the decrease in HBG2 expression is associated with the presence of the 4.9kb deletion, we further examined the gRNAs that exhibited a significant reduction in HBG2 compared to HBG1 expression in ABE (gRNAs 4, 10, 11, 15, 32, 34) and CBE (gRNAs 2, 10, 11, 32, 34, 42) edited cells using quantitative reverse transcription PCR (qRT-PCR). Interestingly, the frequency of the 4.9kb deletion ranged from 2% to 32% in ABE edited cells and from 0% to 12% in CBE edited cells (**Fig 4.1.10e and Fig 4.1.10k**).

To determine the correlation between the reduction in HBG2 chain expression and the frequency of large deletions, Pearson correlation analysis was performed for the aforementioned gRNAs in ABE and CBE edited cells. We observed a strong correlation (r=0.71) in the case of ABE, indicating that the reduction in G gamma chain expression

is primarily attributed to a higher frequency of deletions. In contrast, a lower correlation (r=0.26) was observed for CBE edited cells, suggesting that the decrease in G gamma chain expression is independent of larger deletions and may be influenced by other factors such as biased gamma-globin expression (**Fig 4.1.10f and Fig 4.1.10l**).

Notably, we observed substantial variation in deletion rates among different gRNAs as well as between the two base editors (**Fig 4.1.10e and Fig 4.1.10k**). The differences in DNA sequence composition, which often affect editing efficiency, may contribute to the varied deletion frequencies observed across gRNAs targeting the HBG promoter. Additionally, the processivity of the base editors could account for the differences in deletion patterns observed with the same gRNA when using CBE and ABE. Due to the inability of base editors to bind to an already edited DNA strand, CBE, which achieves a higher editing rate (~50% editing on Day 1), is less likely to interact with the DNA again, thereby reducing the likelihood of deletions compared to ABE, which requires more time to achieve similar editing levels (~50% editing on Day 8) (**Fig 4.1.2b and c**). This observation is further supported by the minimal deletions observed in samples edited with ABE 8e, which has higher processivity (~90% editing within 24 hours), compared to both CBE and ABE 7.10 (**Fig 4.1.11c**) (Richter *et al*., 2020).

Subsequently, we employed high-performance liquid chromatography (HPLC) to assess the levels of hemoglobin tetramers in the differentiated cells, aiming to determine if the increased expression of HBG chains resulted in the production of functional fetal hemoglobin. Remarkably, we observed a significant induction of fetal hemoglobin in all the cells transduced with the top 8 target gRNAs in ABE (gRNAs 2, 3, 4, 10, 11, 15, 32, and 34), while in CBE edited cells, only gRNAs 2, 10, 11, and 42 displayed higher levels of fetal hemoglobin (**Fig 4.1.7e-f**). Consistent with the globin chain analysis, the increase in fetal hemoglobin variants was accompanied by a compensatory downregulation of adult hemoglobin levels in the edited cells.

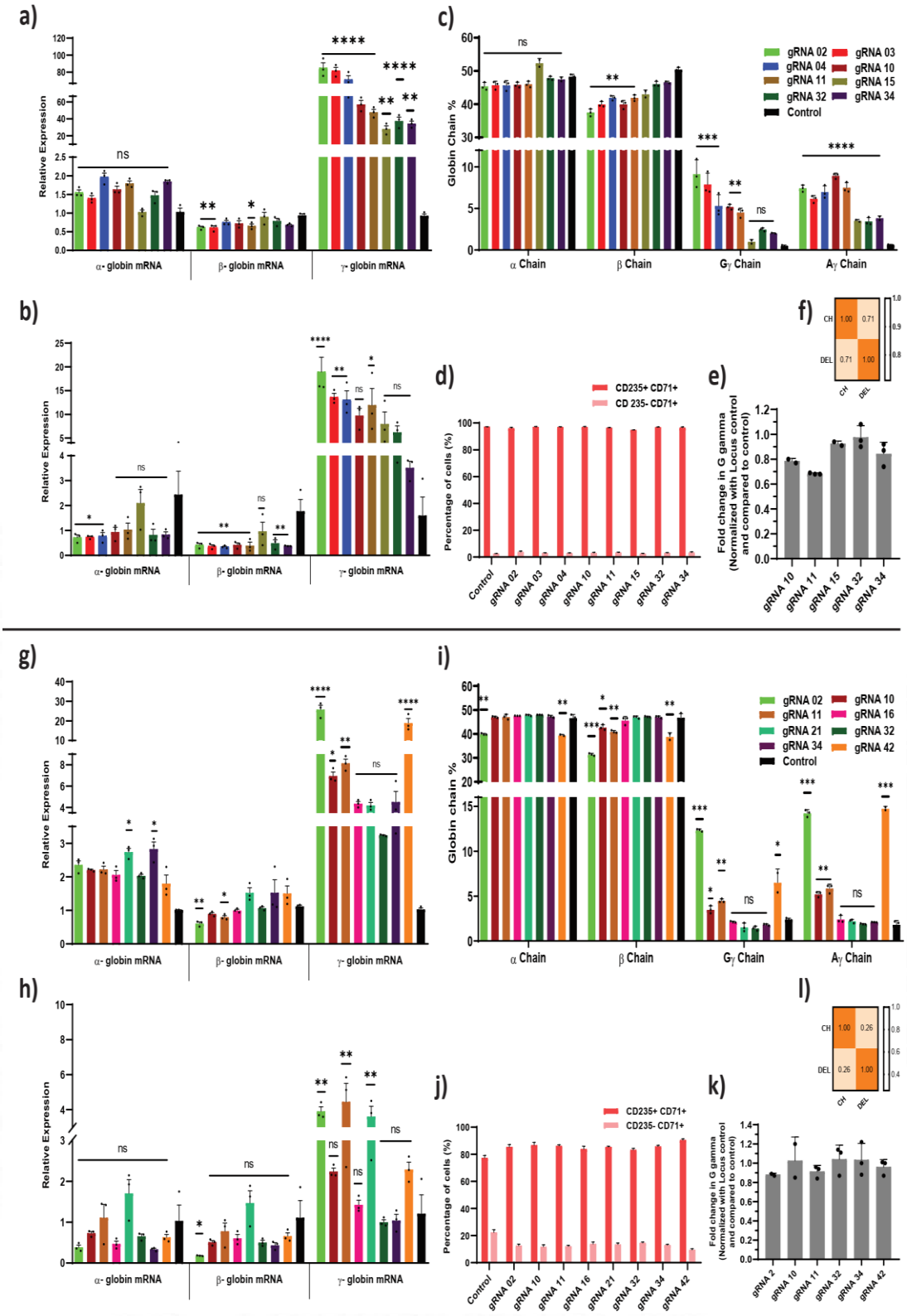


Figure 4.1.10: Manipulation of Adenine and Cytosine Bases in HBG Promoter Influences mRNA and Protein Expression of Globin Chains. Quantification of the relative expression of globin transcripts was performed in samples subjected to adenine and cytosine base editing, both before (a and g) and after (b and h) undergoing erythroid differentiation, using quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR). Following erythroid differentiation, the globin chain composition was

analyzed through Reverse Phase High-Performance Liquid Chromatography (RP-HPLC) (c and i). Flow Cytometry (FACS) analysis was conducted to assess the expression of CD235a and CD71 in edited erythroblasts derived from HUDEP-2 cells (d and j). The presence of a 4.9 kb deletion in HBG2 was examined using qRT-PCR in HUDEP-2 cells edited with an adenine base editor (ABE) (e and k). ABE and CBE (Cytosine Base Editor) edited samples that exhibited a significant reduction in the levels of G gamma chain compared to A gamma chain levels in RP-HPLC were subjected to further analysis for larger deletions using qRT-PCR. A correlation analysis (f and l) was performed between the deletion percentage (DEL) obtained through qRT-PCR and the difference in A and G gamma percentages obtained from HPLC chains. Levels of statistical significance were indicated by asterisks (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). The data presented represents the mean ± SEM from three biological replicates.

To evaluate the relationship between editing efficiency and HbF expression, we examined all the top-scoring gRNAs in ABE and CBE edited cells (**Fig 4.1.7i**). Among the validated top 8 gRNAs in ABE and CBE, gRNAs 2, 10, and 11 with ABE and gRNA 2 with CBE exhibited high target editing efficiency, leading to a corresponding increase in HbF expression. However, gRNA 42 with CBE resulted in only a modest elevation of HbF levels despite its higher editing efficiency. Conversely, gRNAs 3 and 4 with ABE and gRNAs 10 and 11 with CBE demonstrated higher levels of HbF elevation despite their lower base conversion efficiency. The greater number of HbF-positive cells with minimal base editing could be attributed to heterogeneous editing at the target site per cell, considering that there are two copies each of *HBG1* and *HBG2*. Furthermore, regardless of editing at the target region, the binding of the CRISPR-Cas9 complex through the gRNAs at the HBG promoter may disrupt the binding of major transcriptional repressors involved in globin expression (Richter *et al*., 2020).

Among the samples that resulted in higher HbF induction despite lower editing efficiency, we validated gRNAs 3 and 11 using a hyperactive variant of ABE (ABE 8e) to explore whether further elevation in HbF levels could be achieved by increasing the editing efficiency. HUDEP-2 cells stably expressing ABE 8e were transduced with gRNAs 3 and 11, yielding a vector copy number (VCN) of 0.28% for the editor and 0.56% for the gRNA (**Fig 4.1.11e**). We observed a high percentage of base substitution

at the target site (over 95%), accompanied by an increase in HbF-positive cells and gamma globin chains for both gRNAs (**Fig 4.1.11a-b &c**). The erythroid differentiation capacity of the edited cells was equivalent to that of the control (**Fig 4.1.11b**). Moreover, the frequency of larger deletions was significantly reduced, possibly due to the higher processivity rate of ABE 8e (**Fig 4.1.11d**). Consequently, the ABE8e variant improved the base editing efficiency at the target region, resulting in a higher level of HbF induction with a reduced frequency of larger deletions.

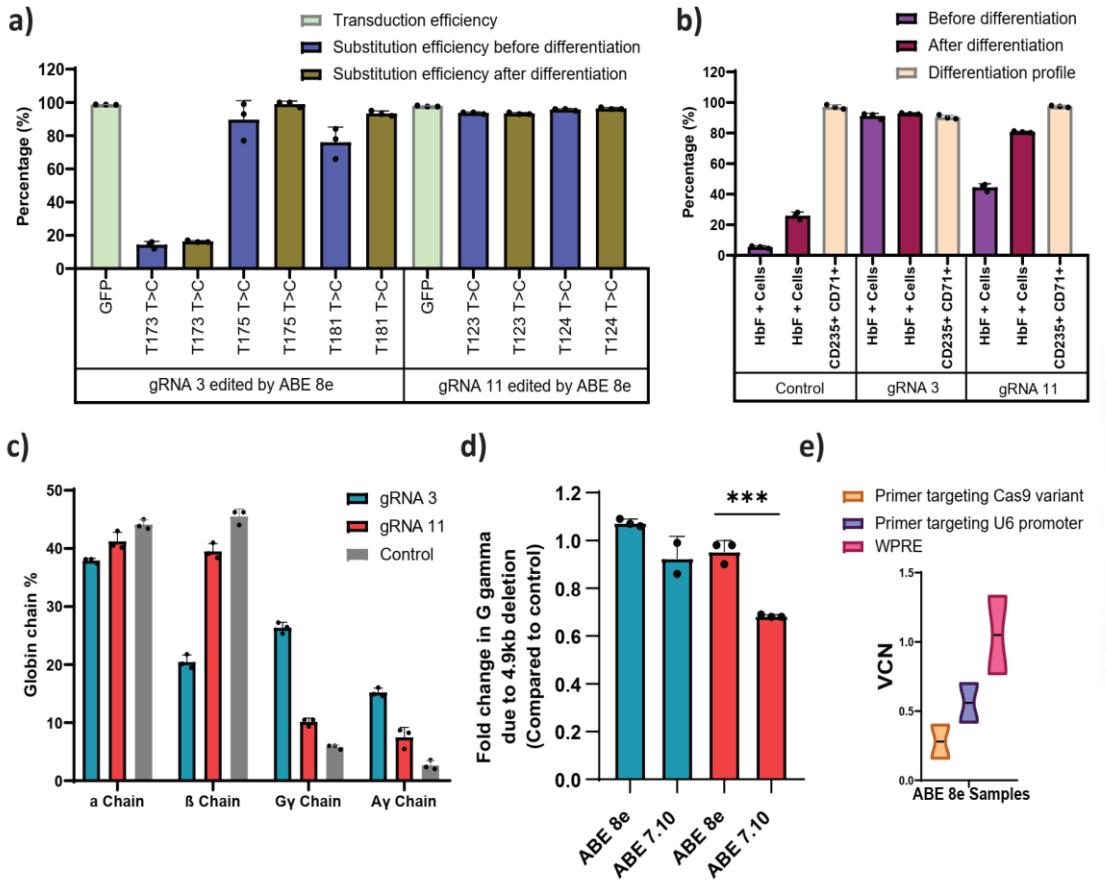


Figure 4.1.11: Assessment of Low-Efficiency Guide RNAs (gRNAs) for Inducing Fetal Hemoglobin (HbF) Expression using Hyperactive Variant Adenine Base Editor (ABE)8e. (a) The transduction efficiency and percentage of individual base conversions were evaluated in HUDEP-2-ABE8e stable cells transduced with gRNA-3 or gRNA-11, both before and after erythroid differentiation. Transduction efficiency was assessed using Flow Cytometry (FACS), while individual base substitutions were analyzed using EditR after Sanger sequencing. (b) Flow cytometry analysis was performed to examine the expression of fetal hemoglobin (HbF) and erythroid maturation markers (CD71+ CD235a+ population) in HUDEP-2-ABE8e stable cells transduced with gRNA-3 or gRNA-11. The percentage of HbF-expressing cells was determined before and after differentiation into erythroblasts. Erythroid differentiation profiles were characterized using CD-235 and CD-71 markers on day 7 of the

differentiation process. (c) Reverse Phase High-Performance Liquid Chromatography (RP-HPLC) analysis was conducted to investigate the composition of globin chains after erythroid differentiation. (d) The presence of a 4.9 kb deletion in HBG2 was examined using quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) in HUDEP-2-ABE8e or ABE7.10 stable cells transduced with gRNA-3 and gRNA-11. (e) The Vector Copy Number (VCN) per cell was determined for the integrated gRNA as well as ABE8e in the corresponding HUDEP-2 stable cell lines. Specific primers targeting Cas9 were used to quantify the gene editor, while primers targeting the U6 promoter were used to quantify the gRNA. The primer targeting WPRE was common for both the gRNA and gene editor. Levels of statistical significance were indicated by asterisks (***$p < 0.001$).

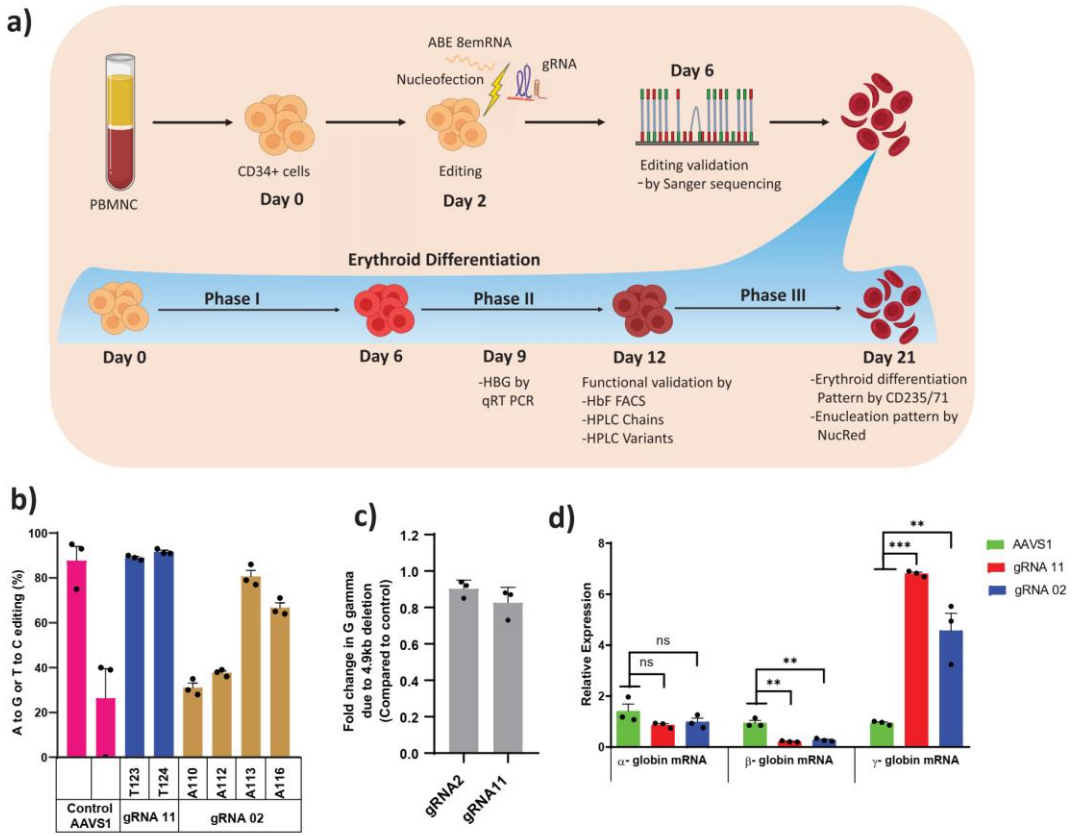
Through our comprehensive screen, we successfully identified several novel individual regulatory regions, as well as validated well-known hereditary persistence of fetal hemoglobin (HPFH) mutations in the proximal promoter of HBG, which play a crucial role in the regulation of gamma-globin expression. Particularly intriguing are gRNA-2 (with ABE or CBE) and gRNA-4 (with ABE), which disrupt the binding site for major gamma-globin repressors, such as BCL11A and LRF/ZBTB7A, while creating binding motifs for transcriptional activators GATA1 and KLF1. This concerted action leads to the overall activation of HBG expression. Base editing at positions -114C>T, -115C>T, and -116A>G disrupts the binding of BCL11A, while base conversions at positions -198T>C and -199T>C affect the binding of LRF/ZBTB7A to the HBG promoter. Furthermore, the installation of -113A>G and -198T>C mutations by gRNA-2 and gRNA-4, respectively, generates binding sites for GATA1 and KLF-1. Notably, the base conversion at position -175T>C induced by gRNA-3 (with ABE) creates a TAL1 binding site. Additionally, the newly discovered gamma-globin regulatory region includes the target base substitutions mediated by gRNA-10, 11, 15, 21, 32, and 34 with ABE or CBE. With gRNAs 10 and 11, ABE converts the nucleotides at positions -123T>C and -124T>C, respectively, while CBE converts the nucleotides at positions -122G>A (on-target editing site) and -117G>A (outside the editing window). Notably, target base substitutions at positions -123T>C and -124T>C result in a greater induction of HBG expression, equivalent to that observed with known HPFH mutations that disrupt the binding of BCL11A. Collectively, the potential gRNAs with high induction

of HbF expression include gRNA-2, 3, 4, 10, and 11 with ABE, as well as gRNA-2 with CBE.

4.1.4 Base editing targeting the -123 region of the HBG promoter in human CD34+ hematopoietic stem and progenitor cells (HSPCs).

To evaluate the therapeutic potential of newly identified targets for inducing gamma-globin expression, we performed base editing on CD34+ hematopoietic stem and progenitor cells (HSPCs) obtained from healthy donors (**Fig 4.1.12a**). Electroporation of ABE8e mRNA along with gRNA targeting the BCL11A binding site (gRNA-2) and the novel -123 cluster (gRNA-11) resulted in highly efficient base editing at the intended site. With gRNA-2, the editing efficiency at specific base positions was observed at -110 (31%), -112 (37%), -113 (80%), and -116 (66%), while with gRNA-11, it was at -123 (89%) and -124 (91%) (**Fig 4.1.12b**). Notably, base editing with gRNA-11 produced a significant number of -123 and -124 mutations at the target site. Subsequent erythroid differentiation of the base-edited CD34+ HSPCs showed significantly elevated levels of HBG expression, with more than 6-fold induction in gRNA-11 and more than 5-fold induction in gRNA-2 compared to the control (AAVS1 edited sample) as determined by qRT-PCR (**Fig 4.1.12d**). Conversely, there was a significant downregulation of HBB expression and no significant change in HBA expression in both target groups. Moreover, the erythroblasts derived from the base-edited CD34+ HSPCs exhibited a substantial increase in HbF protein expression. Flow cytometry analysis and HPLC variant analysis confirmed a robust increase in the proportion of HbF-positive cells and their HbF content in all tested targets, with a greater effect observed in gRNA-11 (**Fig 4.1.12e and g**). Globin chain analysis revealed increased expression of HBG1 and HBG2 globin chains along with reduced HBB globin chain levels (**Fig 4.1.12f**). Importantly, base editing of the HBG proximal promoter using gRNA-2 or gRNA-11 did not impact the enucleation potential or the expression

of erythroid maturation markers CD235a or CD71 (**Fig 4.1.12h-i**). Additionally, the frequency of the 4.9-Kb deletion in CD34+ HSPCs electroporated with ABE8e and gRNA-2 or gRNA-11 was found to be minimal, possibly due to the high processivity and transient expression of the base editor mRNA (**Fig 4.1.12c**). These findings suggest that the level of HbF induction achieved by introducing novel -123 cluster HPFH-like mutations (via gRNA-11) is comparable to naturally occurring -115 cluster HPFH mutations (via gRNA-2) that disrupt the BCL11A binding site. Collectively, our results demonstrate that adenine base editing of the HBG1 and HBG2 promoters to recreate novel -123 cluster HPFH-like mutations holds promise as a therapeutic approach for inducing fetal globin levels and treating beta hemoglobinopathies.



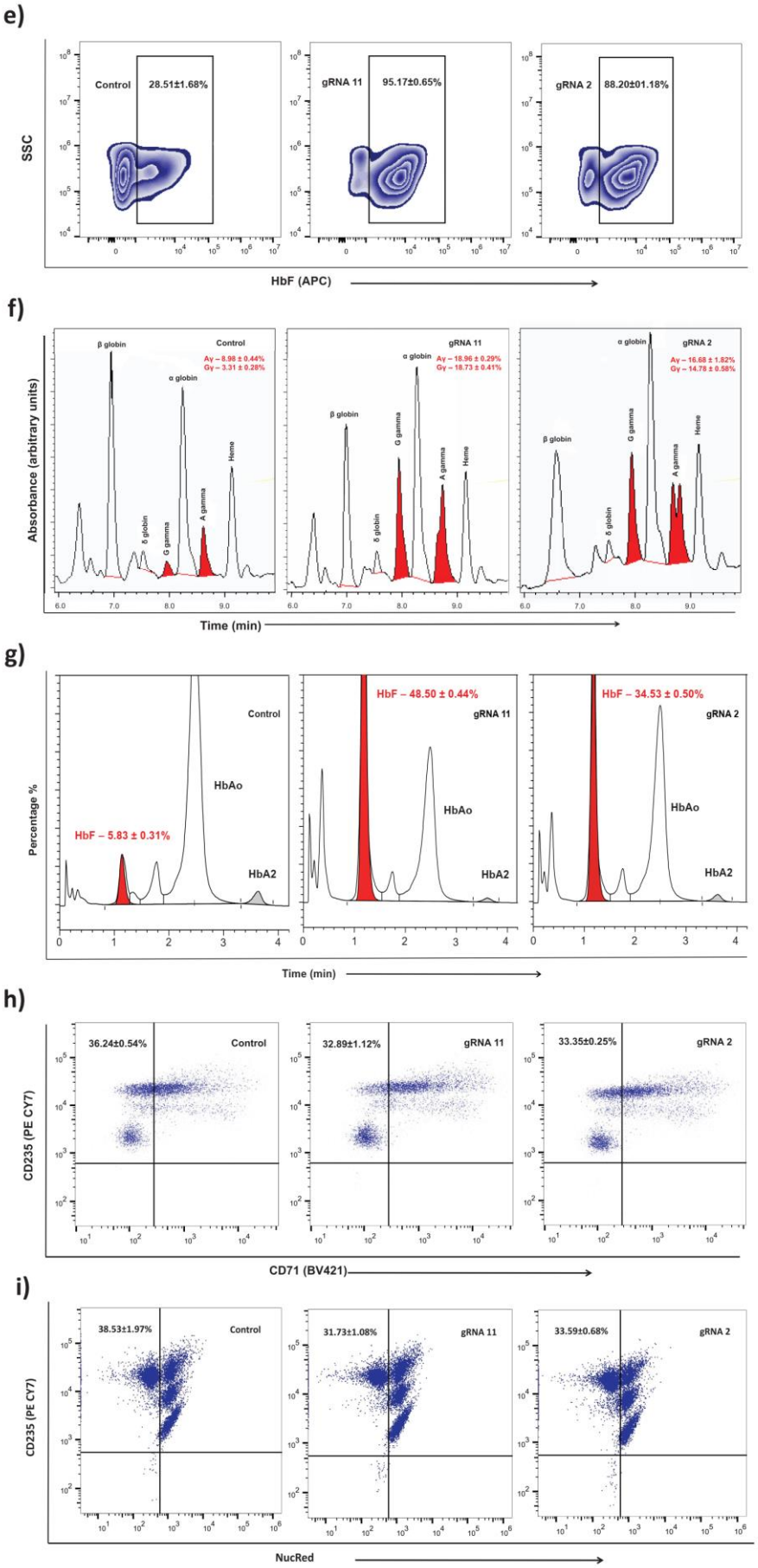

e)
Control
28.51±1.68%
gRNA 11
95.17±0.65%
gRNA 2
88.20±01.18%
SSC
HbF (APC)
f)
Control
gRNA 11
gRNA 2
β globin
α globin
δ globin
G gamma
A gamma
Heme
Absorbance (arbitrary units)
Time (min)
g)
Control
HbF – 5.83 ± 0.31%
HbAo
HbA2
HbF – 48.50 ± 0.44%
gRNA 11
HbF – 34.53 ± 0.50%
gRNA 2
Percentage %
Time (min)
h)
36.24±0.54%
Control
32.89±1.12%
gRNA 11
33.35±0.25%
gRNA 2
CD235 (PE CY7)
CD71 (BV421)
i)
38.53±1.97%
Control
31.73±1.08%
gRNA 11
33.59±0.68%
gRNA 2
CD235 (PE CY7)
NucRed

Figure 4.1.12: Induction of Therapeutic Levels of Fetal Hemoglobin (HbF) in Erythroblasts Derived from Healthy Donor CD34+ Hematopoietic Stem and Progenitor Cells (HSPCs) via Base Editing of the HBG Promoter. (a) Illustration depicting the sequential steps involved in the base editing process of CD34+ hematopoietic stem and progenitor cells (HSPCs). Mobilized CD34+ HSPCs from healthy donors were nucleofected using the MaxCyte system with adenine base editor (ABE)8e mRNA and respective guide RNAs (gRNAs) on the second day of expansion. During the expansion phase, CD34+ HSPCs were analyzed on the sixth day to evaluate the editing efficiency and the presence of a 4.9 kb deletion. (b) Assessment of the efficiency of individual base conversion at the target sites using EditR software following Sanger sequencing. (c) Quantitative analysis of HBG2 deletion (resulting from a large 4.9 kb deletion) by qRT-PCR. The base-edited CD34+ HSPCs were then subjected to a three-phase liquid culture system to promote erythroid differentiation and enucleation. (d) Evaluation of relative expression levels of globin transcripts by qRT-PCR (ΔΔCT) in erythroblasts derived from base-edited CD34+ HSPCs on the ninth day of differentiation. The functional validation of elevated HbF levels was conducted through flow cytometry analysis, high-performance liquid chromatography (HPLC), and reverse-phase HPLC (RP-HPLC) of erythroblasts derived from the indicated samples on the twelfth day of erythroid differentiation. (e) Visualization of HbF-positive cells analyzed by flow cytometry presented as zebra plots. (f) RP-HPLC chromatogram profiles illustrating individual globin chain composition. (g) HPLC chromatogram profile depicting the variants of hemoglobin. (h) Flow cytometry analysis of the expression of erythroid maturation markers CD235a+ and CD71+. (i) Determination of the enucleation pattern in erythroid cells derived from CD34+ HSPCs using flow cytometry analysis with CD235a and NucRed staining. Asterisks indicate statistically significant levels $**p < 0.01$, $***p < 0.001$.

4.1.5 Assessment of Long-term Persistence of editing, Engraftment Potential, and Multilineage Differentiation Capacity of Base-Edited Cells in Mouse Model.

Following the successful standardization of base editing in CD34+ cells, NBSGW mouse models were utilized to investigate the persistence of editing ex vivo, engraftment potential, and multilineage differentiation capacity (McIntosh *et al*., 2015). In the mouse studies, CCR5 was used as the negative control instead of AAVS1, which was employed in the initial cell studies. Conversely, gRNA 2 served as the positive control, while gRNA 11 was the main focus of investigation. To assess the efficiency of editing, DNA was isolated from bone marrow cells obtained from the mice after 16 weeks, and Sanger sequencing was conducted. Notably, the edited cells exhibited augmented persistence in the engrafted cells compared to their *In Vitro* counterparts, as evidenced by a higher proportion of edited cells observed ex vivo (**Fig 4.1.13a**).

Engraftment potential was evaluated by staining bone marrow, spleen, and peripheral blood samples with hCD45 and mCD45 antibodies. The engraftment

percentage was determined by calculating hCD45/(hCD45+mCD45), revealing an engraftment level above 80% in the bone marrow for CCR5 and gRNA 11, while it was approximately 60% for gRNA 2 (**Fig 4.1.13b**). Spleen and peripheral blood samples obtained from the mice exhibited human chimerism ranging from 10% to 20% (**Fig 4.1.13b**).

To investigate the expression of fetal hemoglobin (HbF) in the edited cells, bone marrow cells were fixed, permeabilized, and subjected to staining with CD235 and HbF antibodies. Approximately 5-10% of the cells expressed CD235. CD235-positive cells were subsequently analyzed for HbF expression. Approximately 80% of the cells expressed HbF in study sample, while the positive control exhibited only 70% HbF-positive cells, indicating successful expression of HbF in the edited cells (**Fig 4.1.13b**).

After confirming the persistence of editing and engraftment potential, the cells were further examined for multilineage differentiation capacity. Immunophenotyping was conducted to evaluate lymphoid, natural killer cells, and myeloid lineages. As anticipated, high expression of B cells (CD19) was observed in all samples, followed by moderate expression of the myeloid lineage, suggesting unaffected multilineage differentiation despite the editing process (**Fig 4.1.13c-d**). Furthermore, the long-term population of cells in the bone marrow was analyzed using CD133 and CD34 markers. Approximately 5% of CD34+ cells were detected in the CCR5 sample, while in the gRNA 2 and gRNA 11 samples, the percentages slightly decreased to 2% and 4%, respectively (**Fig 4.1.13f**).

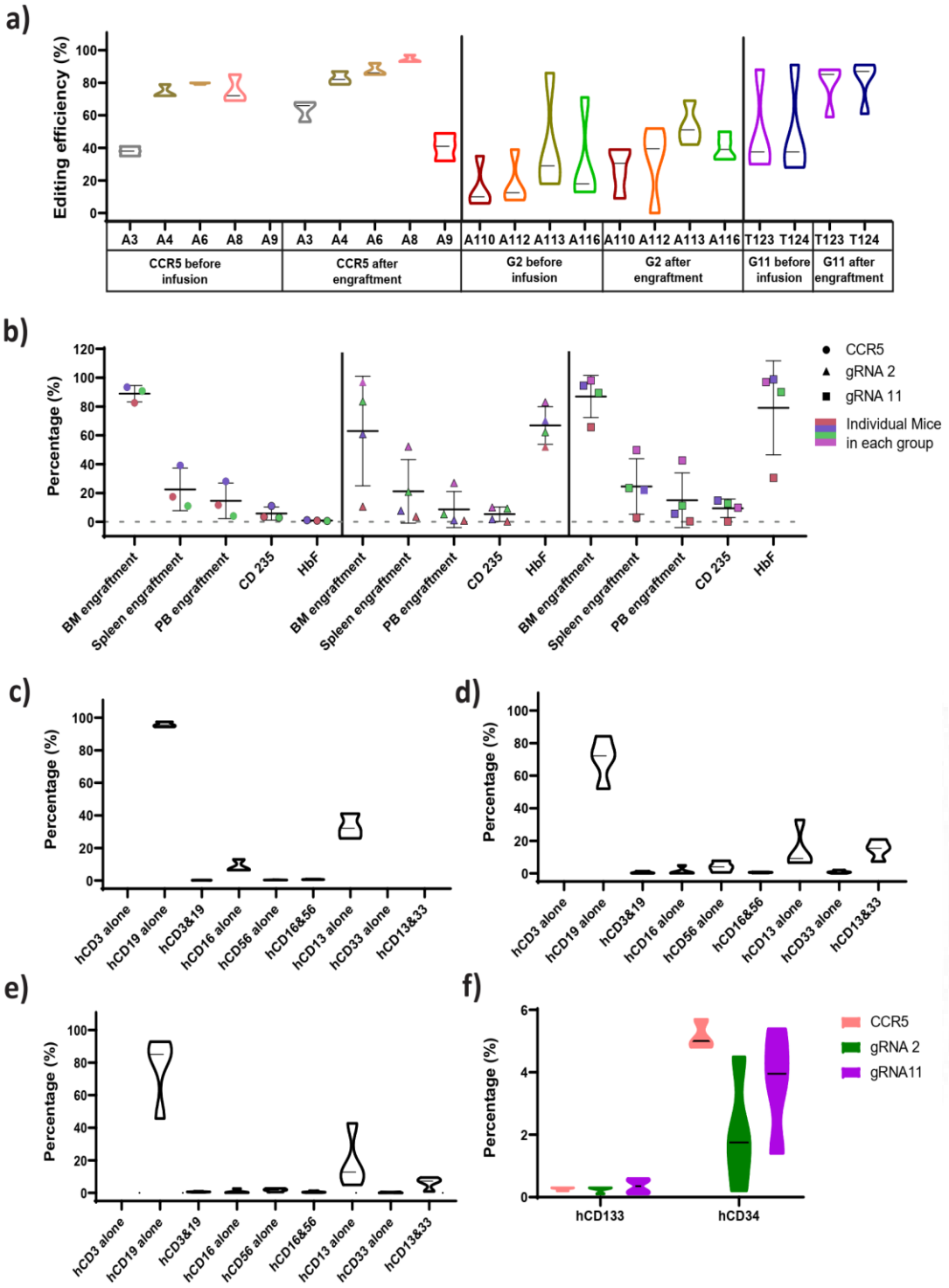


Figure 4.1.13: Evaluation of the Durability of Editing, Engraftment Potential, and Multilineage Differentiation Capacity in Base-Edited Cells using a Mouse Model. (a) Prior to the infusion of edited cells and following their engraftment, the editing efficiency was evaluated by quantifying the relative abundance of successfully edited cells within the overall cell population. (b) The engraftment potential of edited human cells in bone marrow, spleen, and peripheral blood of mice was evaluated. Additionally, the expression of fetal hemoglobin (HbF) in engrafted human CD235 cells was analyzed. (c-d) The multilineage differentiation potential of CCR5, gRNA 2 and gRNA 11-edited cells were examined by

analyzing lineage-specific markers associated with lymphoid cells, natural killer cells, and myeloid cells. (e) The percentage of long-term human repopulating cells within the bone marrow was determined by quantifying the presence of CD133 and CD34 markers.

4.1.6 De Novo KLF1 Binding Site Generated by -123 T>C and -124 T>C HPFH-Like Mutation

To explore the impact of novel HPFH mutations introduced by base editing on transcriptional regulation, we conducted experiments focusing on the -123 T>C and -124 T>C sites. Remarkably, we found that base editing with a single gRNA at these sites resulted in the creation of a consensus binding site for the erythroid transcription factor KLF1 (**Fig 4.1.14 a-b**). This was confirmed through Electrophoretic Mobility Shift Assay (EMSA), where we observed specific binding of KLF1 to the -123 T>C and -124 T>C mutated probe (**Fig 4.1.14 c-d and Fig 4.1.15 a-b**). The combination of these two mutations was crucial for KLF1 binding to the HBG promoter. Subsequently, we performed Chromatin Immunoprecipitation (ChIP) experiments to investigate the direct interaction between KLF1 and the mutated region of the HBG promoter in vivo. Although we detected a weak increase in KLF1 binding to the HBG promoters in cells edited with the -123 and -124 mutations, the effect was modest and similar to the enhancement observed at a negative control locus (**Fig 4.1.15c**). Further investigations are required to confirm whether KLF1 binding to this site is the primary in vivo mechanism driving the up-regulation of gamma-globin in response to the -123 T>C and -124 T>C HPFH mutations.

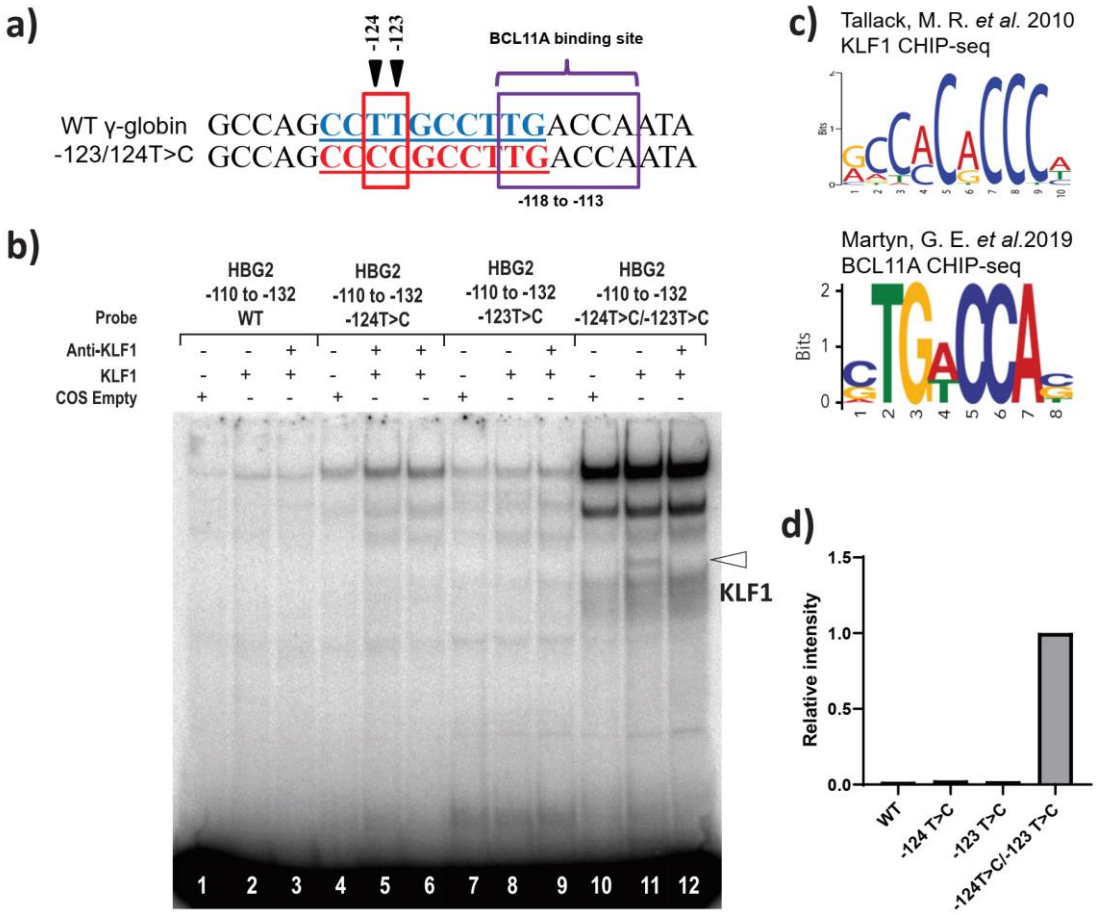


Figure 4.1.14: KLF1 exhibits binding affinity to the −123T > C and -124T > C region of the HBG proximal promoter in vitro. (a) Introduction of a T-to-C mutation at positions –123 and –124 of the HBG promoter (−132 to −110 bp) creates a de novo binding site for KLF1. The wild-type and novel KLF binding motifs are highlighted in blue and red, respectively. (b) In vivo binding motifs of transcription factors KLF1 and BCL11A, as determined by ChIP-Seq in previous studies. (c) Electrophoretic mobility shift assay (EMSA) demonstrates KLF1 binding to the –123T > C/–124T > C probe while failing to bind to the –124T > C probe, –123T > C probe, and wild-type (WT) probe containing the –123T/–124T region of the HBG promoter in vitro. Lanes 1, 4, 7, and 10 represent nuclear extracts from COS cells transfected with an empty pcDNA3 vector. Lanes 2–3, 5–6, 8–9, and 11–12 contain nuclear extracts from COS cells overexpressing KLF1. The binding of KLF1 to the –123T > C/–124T > C HPFH mutant probe is evident in lane 11, with a super shift observed in the presence of anti-KLF1 antibody in lane 12. (d) Quantification of relative band intensities representing KLF1 binding to the probe, determined using Image Lab 6.0.1 (Bio-Rad) software.

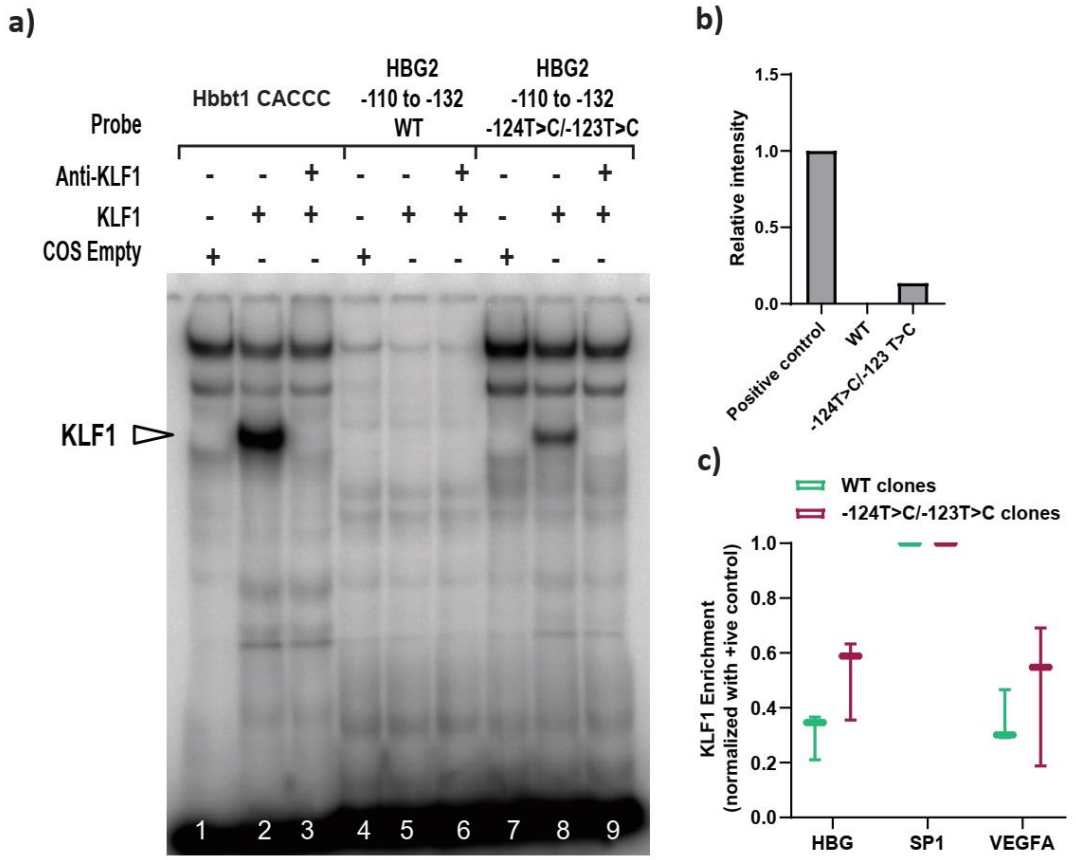


Figure 4.1.15: The recruitment of KLF1 to the site at −123 bp of the HBG proximal promoter was analyzed using electrophoretic mobility shift assay (EMSA) and chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR). (a) EMSA demonstrates the binding of KLF1 to the –123T > C/–124T > C probe, while failing to bind to a wild-type (WT) probe containing the −123/–124 region of the HBG promoter in vitro. Lanes 1–3 represent the Hbbt1-CACCC positive control, lanes 4–6 contain the WT probe for the −123, –124 site (−132 to –110 bp), and lanes 7–9 contain the hereditary persistence of fetal hemoglobin (HPFH) −123/–124T > C mutant probe. Lanes 1, 4, and 7 contain nuclear extracts from COS cells transfected with an empty pcDNA3 vector. Lanes 2–3, 5–6, and 8–9 contain nuclear extracts from COS cells overexpressing KLF1. Binding of KLF1 to the −123/–124T > C HPFH mutant probe is observed in lane 8, with a super shift of KLF1 in the presence of an anti-KLF1 antibody in lane 9. (b) Quantification of relative band intensities from the EMSA using Image Lab 6.0.1 (Bio-Rad) software. (c) ChIP-qPCR analysis of KLF1 binding in HUDEP-2 wild-type (n = 3) and –123T > C/–124T > C cells (n = 3) to measure the relative enrichment of KLF1 at the indicated genomic loci. The promoters of VEGFA and SP1 were used as the negative and positive controls, respectively. The values were normalized to the positive control locus (SP1). Data are presented as mean ± SEM from three biological replicates. Asterisks indicate levels of statistical significance (*p < 0.05).

4.1.7 Analysis of off-target effects and gene expression changes following base editing at the HBG promoter:

Adenine base editors (ABEs) and cytosine base editors (CBEs) have been reported to generate Cas-dependent DNA off-target effects and transient Cas-

independent RNA off-target effects at low levels. However, Cas-independent DNA off-target effects have been shown to be minimal and undetectable (Andrew V. Anzalone, 2020). To assess the potential Cas-dependent DNA off-target effects, we utilized the Cas-OFFinder tool to predict potential off-target sites for the novel gRNA (gRNA11). The identified target regions were subjected to deep sequencing using next-generation sequencing (NGS). Despite the high efficiency of on-target editing, we did not observe off-target editing at the predicted target sites (**Fig 4.1.16a**).

Furthermore, to investigate whether base editing induces significant spurious RNA deamination across the transcriptome, we performed transcriptome-wide RNA sequencing on stable cell lines treated with ABE or CBE, with or without gRNA-2 and gRNA-11. The frequency distribution of A-to-I conversion (in ABE-treated cells) or C-to-U conversion (in CBE-treated cells) in the base-edited samples was found to be highly similar to that of the parental stable cell line (**Fig 4.1.16b-e**).

To further confirm that editing the gamma-globin promoter does not impact the expression of other genes involved in globin regulation, we conducted a differential analysis on these samples specifically targeting genes associated with globin regulation. Our analysis revealed no significant differences between the edited cells and control cells, except for the gamma and delta-globin genes (**Fig 4.1.17**).

Taken together, these findings support the utility of ABE and CBE in generating precise point mutations in the homologous HBG1 and HBG2 promoters. This approach has the potential to increase the number of HbF-positive cells and overall HbF production, while significantly reducing the frequency of larger deletions.

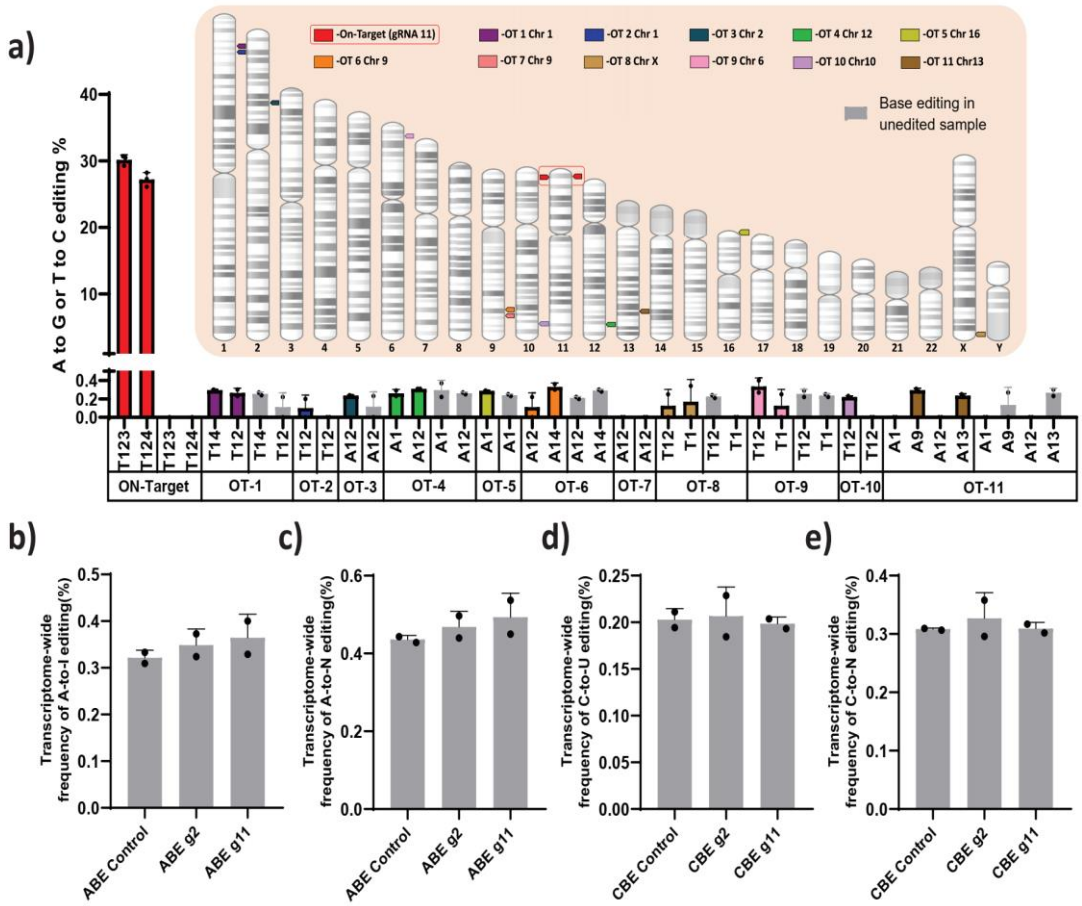


Figure 4.1.16: Assessment of Cas-dependent DNA off-target and Cas-independent RNA off-target editing by adenine base editor: (a) Analysis of base conversions at the top 11 Cas-dependent DNA off-target sites in adenine base editor (ABE) 7.10-stable cells edited with guide RNA (gRNA)-11, alongside the on-target events. The chromosome positions of the off-target and on-target loci are indicated. (b-e) Representation of the frequency of transcriptome-wide cellular levels of A-to-I (b), A-to-N (c), C-to-U (d), and C-to-N (e) RNA editing in BE3-stable cells (cytosine base editor, CBE), ABE 7.10-stable cells (ABE), BE3-stable cells edited with gRNAs-2 or gRNAs-11, and ABE-edited cells with gRNAs-2 or gRNAs-11. The data shown are the mean ± standard deviation (SD) of two technical replicates.

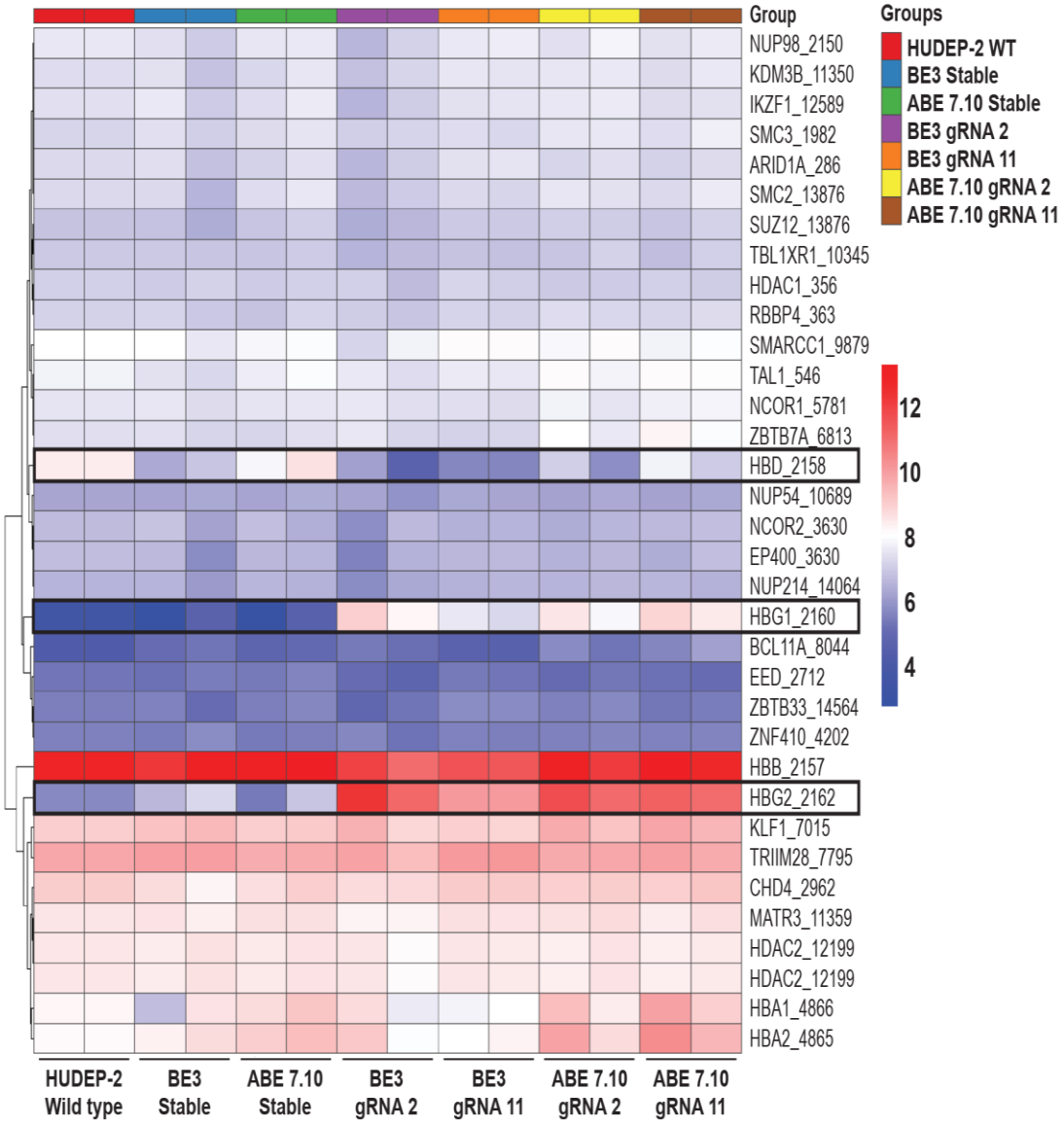


Figure 4.1.17: Expression profile of HBG regulators following base editing: The comparative analysis of differential gene expression was performed on a selection of 34 genes associated with the regulation of gamma-globin. The expression levels of these genes were assessed across various samples including unedited HUDEP-2 wild type (WT), cytosine base editor (CBE) control, adenine base editor (ABE) control, edited ABE (with guide RNA [gRNA]-2 or -11), and edited CBE (with gRNA-2 or -11). The gene expression values are depicted in the form of a heat map, illustrating the relative expression patterns.

Section 4.2: Monoallelic integration of Factor IX transgene in ITGA2B gene and their concerted expression can ameliorate haemophilia B

This study introduces a novel gene replacement strategy called MICE (Monoallelic Integration and Concerted Expression). The purpose of this strategy is to achieve coordinated expression of genes involved in a common biological process or pathway. The approach involves inserting the coding sequence of one gene into another gene locus that functions as a partner in the process. In this study, we employed a well-established model to provide evidence and assess the effectiveness of the MICE approach. Specifically, we utilized the ITGA2B promoter to drive the expression of coagulation factors (FIX), resulting in the sequestration of these factors within platelets. Through the genetic engineering of hematopoietic stem and progenitor cells (HSPCs), we generated platelets that can circulate in the bloodstream (Zhang *et al*., 2010). When a vascular injury occurs, these platelets are recruited to the site of injury as the initial stage of hemostasis. Upon activation, the platelets release the necessary factors for secondary hemostasis. One notable advantage of this strategy is its potential application as a therapeutic approach for patients with hemophilia who already have antibodies against the coagulation factors (Franchini and Mannucci, 2011). By optimizing the CRISPR-based Homology-Directed Repair (HDR) method with different donor constructs and dosages, we achieved precise integration of FIX or GFP expression into the ITGA2B locus, resulting in coordinated expression of both genes. To validate the strategy, we initially optimized the HDR approach in HEL cell lines, which naturally express the ITGA2B gene, and then extended the methodology to HSPCs for further validation. The edited HSPCs were differentiated into megakaryocytes in vitro, and platelets derived from these cells were assessed for GFP and FIX expression. Additionally, we confirmed the storage of the gene of interest in the alpha granules of

the megakaryocytes. Lastly, we analyzed the edited cells for any potential off-target effects associated with the guide RNA (gRNA) targeting the ITGA2B locus.

4.2.1 Optimization of homology-directed repair (HDR) for gene editing in the ITGA2B locus:

To achieve HDR, we induced a double-strand break in intron-1 of the ITGA2B locus by introducing a ribonucleoprotein (RNP) complex consisting of the Cas9 protein and a synthetic guide RNA (gRNA) targeting the specific locus of interest. The RNP complex was delivered into cells through electroporation and subsequently transduced with AAV6 carrying a donor template comprising 500bp homology arms, desired regulatory elements, and the gene of interest (**Fig 4.2.1a**). Utilizing the CHOP-CHOP tool (version 3), we identified the top 10 gRNAs targeting intron-1 of the ITGA2B gene. Principal component analysis (PCA) was performed on the editing efficiency and microhomology-mediated end-joining (MMEJ) score obtained from in-silico tools to select the optimal gRNAs (**Fig 4.2.1b & Fig 4.2.2a**). The MMEJ score was evaluated as studies indicate that HDR efficiency is significantly enhanced when the gRNA activates the MMEJ pathway. Through in-silico analysis of editing efficiency and MMEJ score, we identified the top 4 gRNAs. To verify their editing efficiency in vitro, we performed validation experiments in HEL cells with stable Cas9 expression. Lentivirus carrying the gRNA and GFP gene was used to evaluate both editing and transduction efficiency in the HEL stable cells. Among the four gRNAs (gRNA 1, 7, 9, & 10), gRNA 10 exhibited the highest editing efficiency and was chosen for further editing. Notably, all the gRNAs demonstrated a 100% transduction efficiency. (**Fig 4.2.1c**).

HDR was performed using the AAV6-based donor template at two different viral titres (high and low), and knock-in efficiency was evaluated through qualitative and quantitative assays. The donor construct employed was AAV6-Ef1α-GFP, and its

constituents are depicted in a schematic diagram presented in **Figure 4.2.1d**. Based on the titration of AAV6 carrying the donor template (**Fig 4.2.2b**), the multiplicity of infection (MOI) was approximately 130k and 45k for high and low titres, respectively. After 48 hours of transduction, DNA was isolated and knock-in events were qualitatively validated through whole-length PCR (WLP) and in-out PCR (IOP). The regions for primer design targeting WLP and IOP were clearly marked and indicated in the accompanying diagram (**Fig 4.2.1e**). In the case of WLP, the inserted fragment resulted in a band at 2691 bp, while the wild-type band was around 1368 bp. In IOP, a band of approximately 1269 bp was obtained (**Fig 4.2.1f**). The presence of an unspecific band observed between the inserted fragment and the wild-type band in WLP could be attributed to the formation of heteroduplexes between these two entities (**Fig 4.2.2c**).

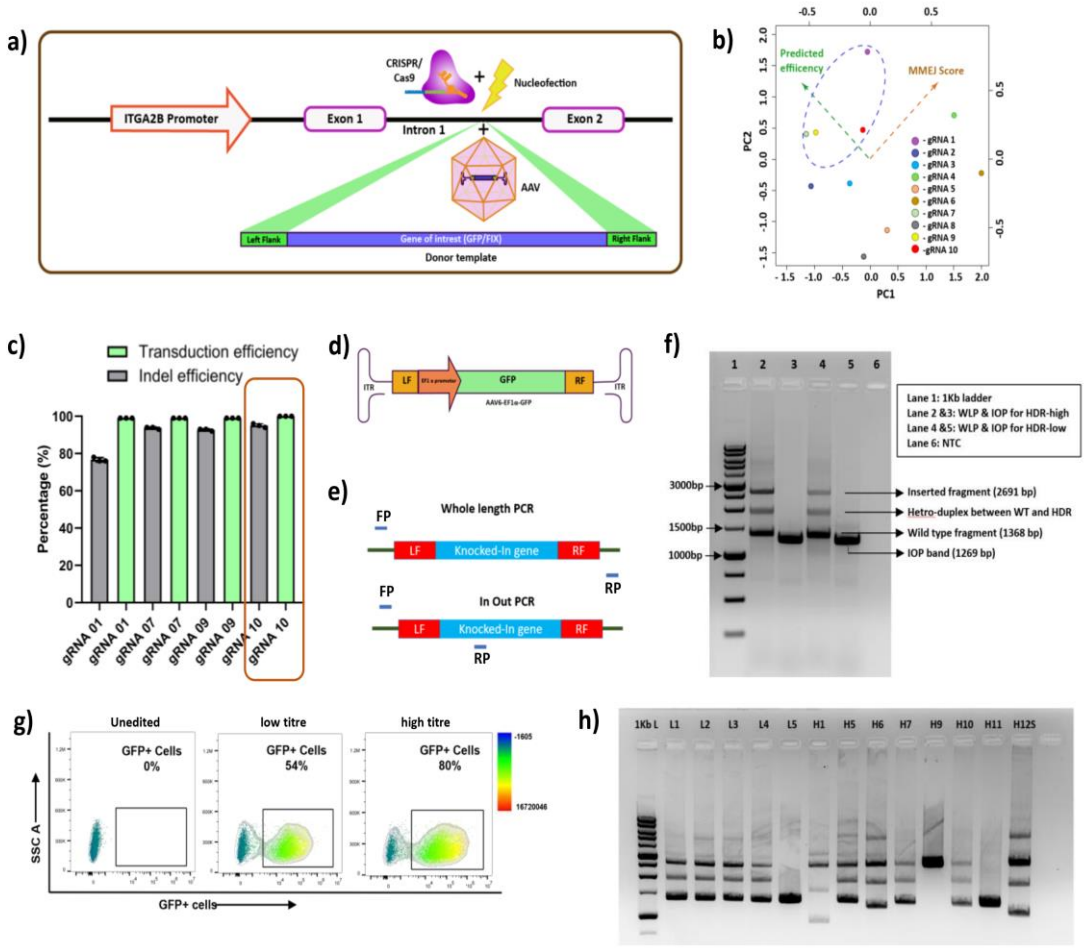


Figure 4.2.1: Enhancement of HDR-mediated gene editing technique at the ITGA2B locus: Insights encompassing gRNA selection, donor template design, and verification of knock-in through molecular and phenotypic assays. a) Schematic illustration of the insertion of the gene of interest within ITGA2B intron-1 using CRISPR/Cas9 and an AAV6-based donor template. b) Principal Component Analysis employed to identify the most efficient gRNA based on computational predictions of editing efficiency and MMEJ score. c) Screening of highly efficient gRNAs in HEL cell line utilizing a lentiviral delivery system, followed by validation of transduction and editing efficiency. d) Development of an AAV6-based donor construct featuring 500 bp Homology Arms (depicted in orange) surrounding the GFP Gene regulated by the EF1α Promoter. e) Schematic representation of primer positioning for Whole Length PCR (WLP) and In-Out PCR (IOP) techniques, enabling evaluation of the success of gene knock-in; FP and Rp denote the forward and reverse primers, respectively. f) Verification of gene integration at the ITGA2B locus via WLP and IOP for high and low titer donor delivery, with results displayed through agarose gel electrophoresis. g) Assessment of knock-in efficiency through flow cytometry for low and high titer donor delivery at the ITGA2B locus. h) Validation of HDR clones edited using low and high titer AAV6 donor through WLP, illustrated through agarose gel electrophoresis. The labels "L" and "H" denote low and high HDR, respectively.

Flow cytometry analysis revealed that the high-titre sample exhibited a knock-in efficiency of approximately 80%, while the low-titre sample had an efficiency of around 54% (**Fig 4.2.1g**). Heat map statistical plots were used to visually represent the increasing GFP expression per individual cell, corresponding to the enhanced efficiency of knock-in events (**Fig 4.2.1g**). Single-cell sorting was performed in both samples to evaluate monoallelic and biallelic integration. All the obtained clones showed varying integration patterns with almost 100% GFP expression, except for clone L2, which exhibited over 80% expression (**Fig 4.2.1h & Fig 4.2.2d**). Among the clones from the low-edited sample, four clones displayed biallelic integration, while one clone did not show any integration despite GFP expression (**Fig 4.2.1h**). The lack of integration in this particular clone may be attributed to the low frequency of random AAV6 integration into the genomic DNA. Out of the eight clones obtained from the high-edited sample, six clones exhibited biallelic integration, of which three clones showed heterogeneous patterns of wild-type bands, indicating that a higher dose of AAV donor may not be ideal for preserving genomic integrity (**Fig 4.2.1h**). Among the remaining two clones, one had monoallelic integration expressing only GFP without any CD41a (ITGA2B) expression (**Fig 4.2.2e**), while the other clone did not exhibit any integration but still

expressed GFP, suggesting random integration (**Fig 4.2.1h**). Comparison of mean fluorescence intensity (MFI) among the clones revealed variations even among clones with biallelic integration (**Fig 4.2.2f**). The clone exhibiting monoallelic expression had the highest MFI, followed by the clones with random integration. Collectively, these findings suggest that targeted integration of the GFP coding sequence at the ITGA2B locus is a feasible approach; however, careful consideration of the optimal AAV donor titre is necessary, as elevated titres may have deleterious genomic effects.

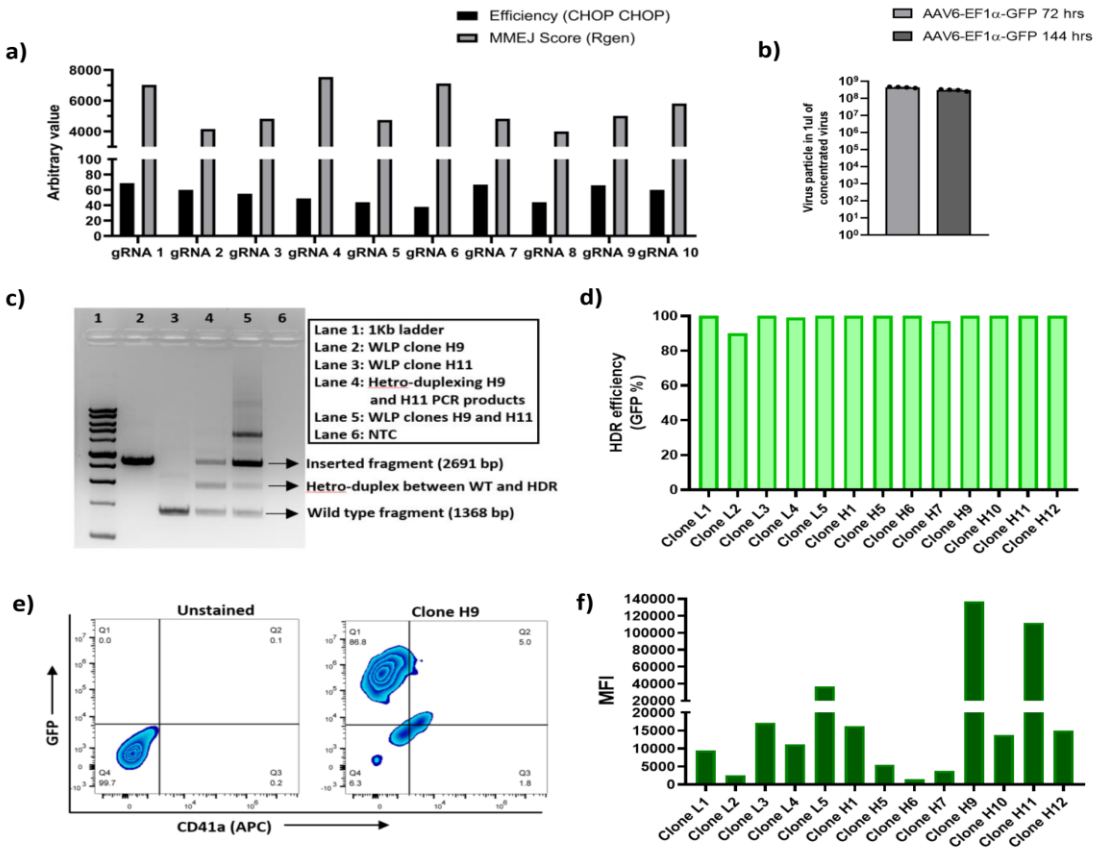


Figure 4.2.2: Assessment of the efficiency of CRISPR-mediated genome editing at the ITGA2B locus and its outcomes: In silico analysis and experimental validation. a) In silico analysis was employed to evaluate the editing efficiency and Microhomology-mediated end joining (MMEJ) scores for the top 10 gRNAs targeting ITGA2B Intron-1. b) The titer value of adeno-associated virus serotype 6 (AAV6) containing the donor template was assessed at 72 hours and 144 hours post-collection. c) Investigation of the underlying cause of the non-specific band observed between the wild type and knock-in bands during Whole Length PCR (WLP). d) The expression of GFP in clones derived from both low and high HDR samples indicated the stable integration of the donor template. The labels "L" and "H" denote low and high HDR, respectively. e) Flow cytometry was used to evaluate the expression of CD41a and GFP in

Clone H9. f) The mean fluorescence intensity (MFI) of GFP expression was quantified in both low and high HDR clones.

4.2.2 Maximizing monoallelic integration efficiency through donor template diversity in HDR-mediated gene editing optimization at the ITGA2B locus:

Following the successful establishment of the CRISPR-based HDR knock-in technique in HEL cells, we focused on optimizing various donor template constructs with different titrations to achieve precise and efficient monoallelic integration of a transgene at the ITGA2B locus. To optimize the titration, we utilized the AAV6-EF1α-GFP donor and observed a concurrent increase in HDR efficiency with an exponential rise in the titre (**Fig 4.2.3a**). When knock-in efficiency surpassed 50%, a significant decrease in CD41a expression at the ITGA2B locus indicated that the transgene was integrated in a biallelic manner (**Fig 4.2.3a**). Validation of the knock-in efficiency was further confirmed using droplet digital PCR (ddPCR), which showed slightly lower values compared to flow cytometry, but the overall trend remained consistent (**Fig 4.2.3b**). Prior to HDR, HEL cells exhibited an indel efficiency of over 80% (**Fig 4.2.4a**). The stability of HDR-mediated integration was assessed by evaluating HDR efficiency over time (**Fig 4.2.4b**). A decrease in expression from day 2 to day 4 suggested the dilution of episomal donor templates by day 4. However, expression levels remained constant from day 4 to day 13 (**Fig 4.2.4b**). In contrast, the sample treated with virus alone (at 128k MOI) demonstrated a decline in GFP expression from 60% to 2%, indicating episomal expression. Notably, the 2% expression level observed in the virus-only sample indicated a rare but expected occurrence of random integration in AAV (**Fig 4.2.4b**). The predominant post-indel patterns present in the knocked-in samples were validated, including a 9bp deletion, 1bp insertion, and 2bp deletion (**Fig 4.2.4c**). A principal component analysis (PCA) was conducted to examine the relationship between

the indel pattern and HDR efficiency, revealing that as HDR increased, the 9bp deletion decreased, suggesting a preference for short deletions in HDR-mediated repair (**Fig 4.2.3c**). Microhomology-mediated end joining (MMEJ) pathway is responsible for the increased indel size, improving HDR efficiency as reported in previous studies (Kosicki *et al*., 2022; Lee *et al*., 2022).

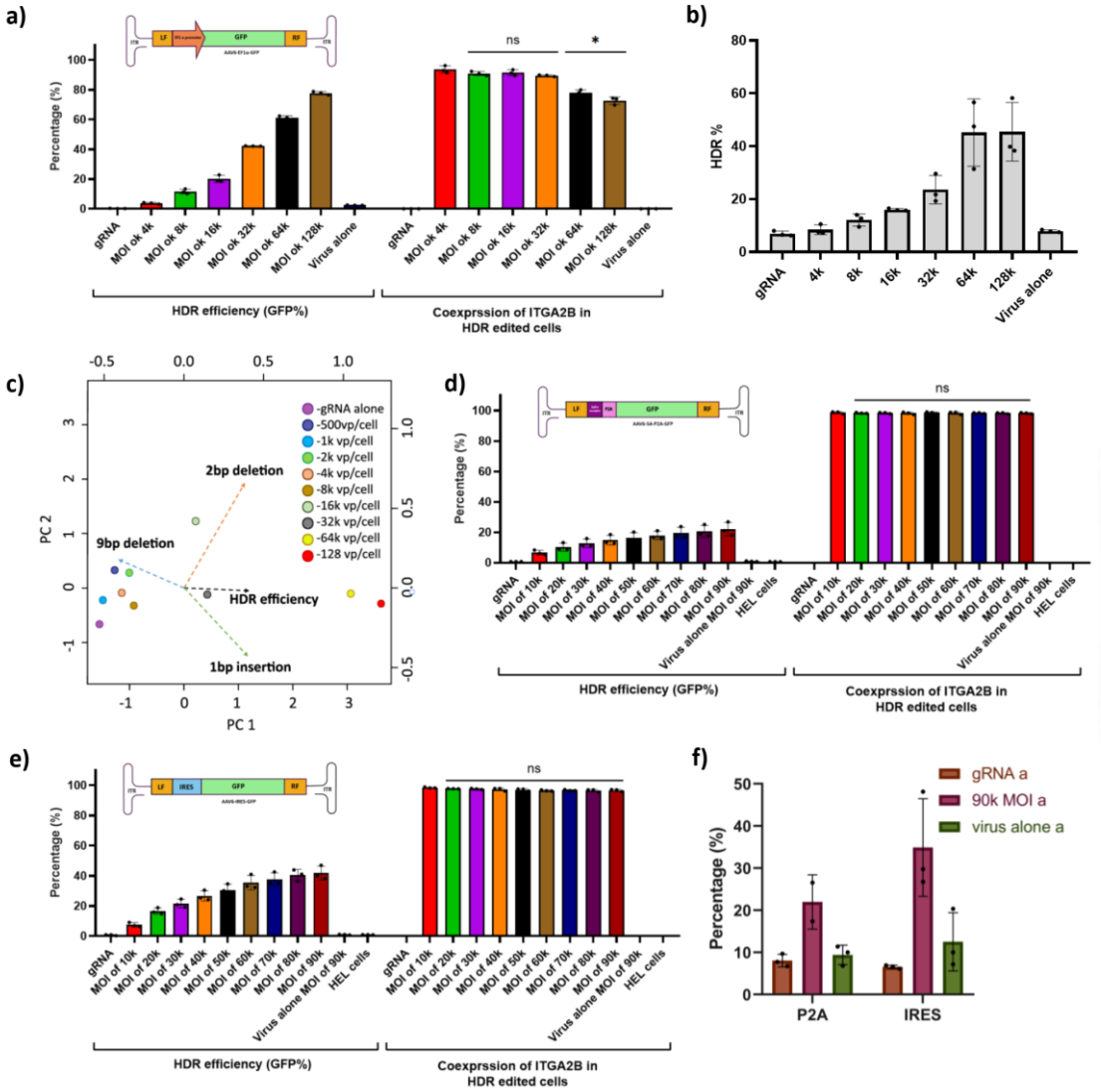


Figure 4.2.3: Enhancement of monoallelic integration efficiency by diversifying the donor template in HDR-mediated gene editing optimization at the ITGA2B locus, and analysis of indel patterns and their correlation with knock-in efficiency. a) HDR was performed at the ITGA2B locus using CRISPR/Cas9 and an AAV6 donor template, consisting of 500bp homology arms flanking GFP under the EF1α Promoter in HEL cell lines. Knock-in efficiency and co-expression of ITGA2B (CD41a) in HDR-

edited cells were assessed using flow cytometry at different multiplicity of infection levels. b) Quantification of knock-in efficiency was achieved through droplet digital PCR (ddPCR) for samples edited with the AAV6-EF1α-GFP donor template. c) Principal Component Analysis (PCA) was used to identify the significant indel pattern favoring HDR by analyzing the indel patterns obtained from HDR samples treated with varying concentrations of the donor template, correlated with HDR efficiency. d & e) Targeted knock-in of two promoterless donor templates, AAV6-SA-P2A-GFP and AAV6-IRES-GFP, at the ITGA2B locus in HEL cell lines, followed by flow cytometry-based analysis to evaluate knock-in efficiency and co-expression of ITGA2B (CD41a) in HDR-edited cells at different multiplicity of infection levels. e) Principal Component Analysis (PCA) was employed to identify the significant indel pattern favoring HDR by analyzing the indel patterns obtained from HDR samples treated with varying concentrations of the donor template, correlated with HDR efficiency. f) ddPCR analysis was performed to quantify knock-in efficiency in HEL cells edited with the promoterless donor templates, AAV6-SA-P2A-GFP and AAV6-IRES-GFP. The data are presented as mean ± SEM from three experimental replicates, and asterisks indicate levels of statistical significance (*$p < 0.05$).

Two promoterless donors, namely AAV6-SA-P2A-GFP and AAV6-IRES-GFP, were utilized to enable transgene expression under the control of the ITGA2B promoter. The titre concentrations of these donors were systematically varied based on data obtained from the previous donor (**Fig 4.2.3d & e**). MOIs ranging from 10k to 90k were employed to prevent biallelic integration of the transgene resulting from HDR efficiencies above 50%. Plateauing of HDR efficiency was observed even with increased titre in the case of AAV6-SA-P2A-GFP, with a maximum efficiency of approximately 20% (**Fig 4.2.3d**). On the other hand, AAV6-IRES-GFP demonstrated improved HDR efficiency, reaching around 40% efficiency when MOIs were above 80k. In both AAV6-SA-P2A-GFP and AAV6-IRES-GFP donors, CD41a expression remained unaffected even at high titres, as HDR efficiency did not exceed 50%. However, the efficiency in both promoterless donors was lower compared to the AAV6-EF1α-GFP donor. The knock-in efficiency of the SA-P2A and IRES constructs was evaluated using ddPCR at MOI 90k, and the pattern obtained was consistent with flow cytometry analysis (**Fig 4.2.3f**). Following knock-in, quantitative reverse transcription PCR (qRT-PCR) was performed to validate GFP mRNA levels, which showed a significant increase compared to the virus-only treated sample (**Fig 4.2.3d**).

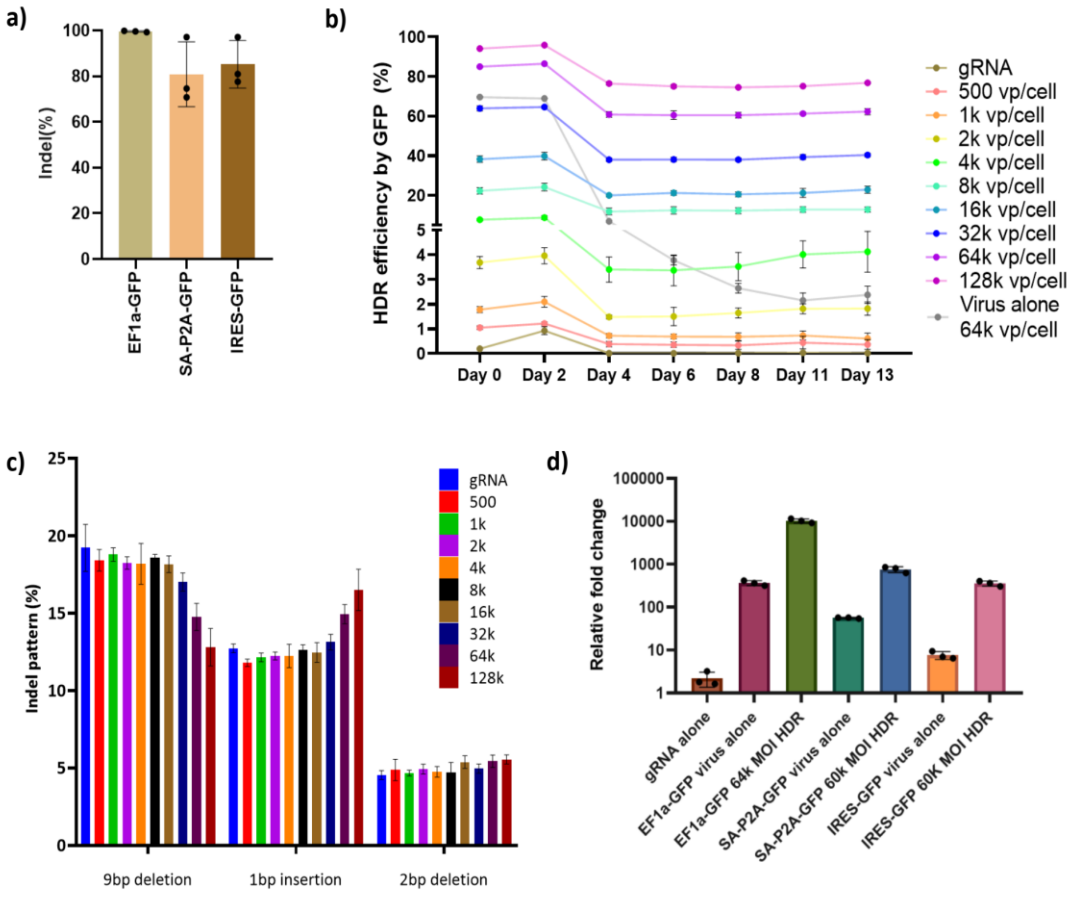


Figure 4.2.4: Evaluation and verification of genome editing strategies for targeted gene knock-in and expression in HEL cells using CRISPR/Cas9 and an AAV6-based donor template. a) Assessment of Indel efficiency at day 4 following HDR editing using gRNA IT10. b) Temporal validation of HDR efficiency through flow cytometry-based analysis of GFP expression in samples edited with the AAV6-EF1α-GFP donor template at various multiplicity of infection levels. c) Identification of the prominent Indel pattern by targeted sequencing of the ITGA2B locus after knock-in using the AAV6-EF1α-GFP donor template. d) Relative quantification of GFP mRNA expression in HDR samples edited with three different donor templates, using quantitative reverse transcription PCR (qRT-PCR).

The detailed methodology employed for conducting the HDR experiment has been illustrated in **Figure 4.2.5a**. Adequate virus titres were successfully achieved for all three donor constructs (**Fig 4.2.5b, c, & d**). The titres of AAV6-EF1α-GFP donor virus peaked at 72 hours post-transfection, while for AAV6-SA-P2A-GFP and AAV6-IRES-GFP donors, the highest titres were obtained at 144 hours post-transfection (**Fig 4.2.5b, c, & d**). The CRISPR-mediated HDR-based knock-in method was effectively established in HEL cells, and different donor template constructs were optimized to achieve monoallelic integration of the transgene at the ITGA2B locus. Biallelic

integration of the transgene occurred when HDR efficiency exceeded 50%, resulting in a significant decrease in CD41a expression. However, since promoterless constructs had HDR efficiencies below 50%, only monoallelic transgene integration was achieved.

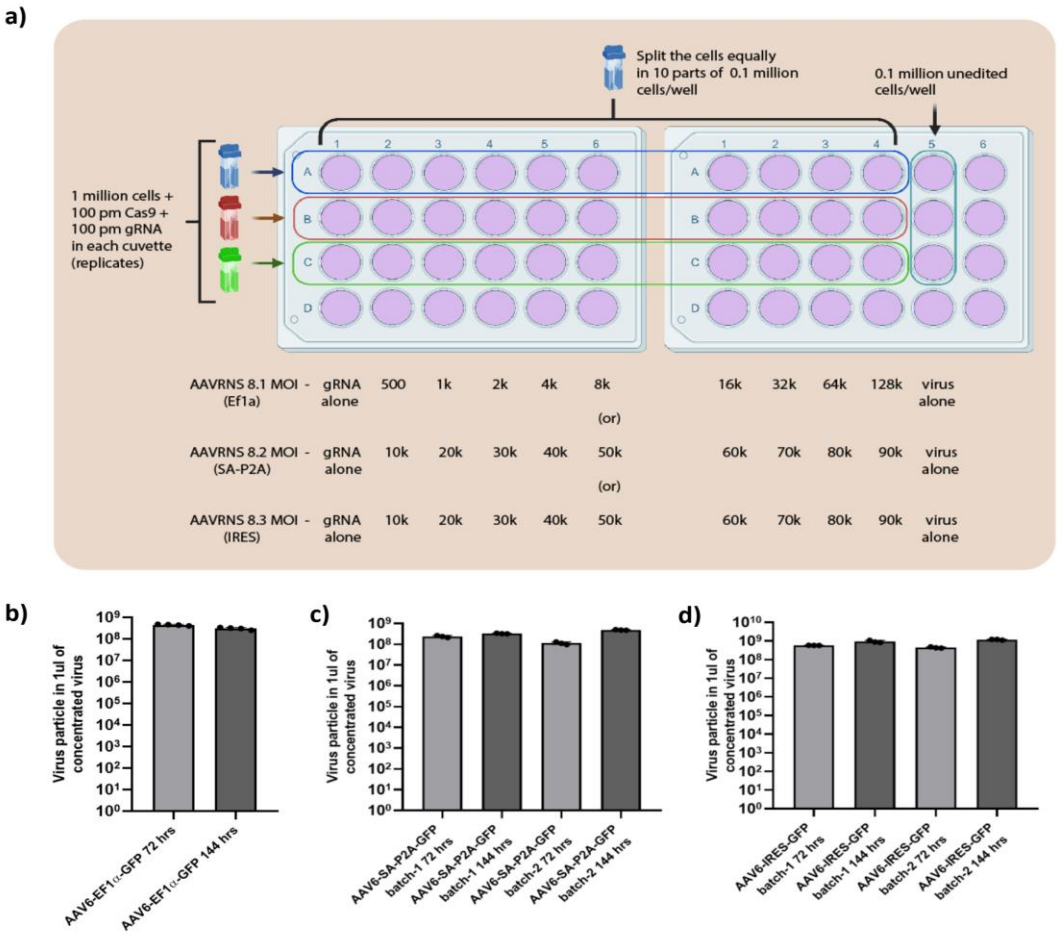


Figure 4.2.5: Methodology and titration of donor templates in HDR-mediated gene editing using AAV6-based donors: a) Elaboration of a comprehensive methodology for conducting HDR in Hel cell lines, incorporating all three donor templates. b-d) Quantitative assessment of AAV6 containing each specific donor template through relative quantification (ΔΔCT) using quantitative reverse transcription PCR (qRT-PCR).

4.2.3 Enhancing the efficiency of HDR-mediated gene knock-in under the ITGA2B promoter while mitigating CD34+ cell toxicity through delivery parameter manipulation:

Following meticulous optimization of the promoterless donor template for expression under the ITGA2B promoter in HEL cells, successful introduction of the template into CD34+ cells was achieved. CD34+ cells possess the ability to differentiate

into various blood cell lineages, including megakaryocytes that harbor an active ITGA2B promoter. These megakaryocytes play a key role in platelet generation and were specifically targeted for transgene expression. Both SA-P2A-GFP and IRES-GFP donors were utilized for HDR to insert GFP under the ITGA2B promoter. After successful HDR, the cells were induced to differentiate into megakaryocytes, resulting in ITGA2B promoter activation and subsequent transgene expression. CD41a (ITGA2B) expression gradually increased during the differentiation process, exceeding 90% on day 11 compared to the initial 60% on day 5 (**Fig 4.2.6a**). Furthermore, mature megakaryocytes co-expressed CD41a and CD42b, with their expression levels also increasing during megakaryocytic differentiation, reaching up to 80% (**Fig 4.2.6a**). In accordance with findings from a previous study, we successfully achieved robust megakaryopoiesis (Perdomo *et al*., 2017). Notably, unedited CD34+ cells and those treated with gRNA alone demonstrated robust megakaryocytic maturation, with 80% CD42b expression on day 11. In contrast, knocked-in samples and samples treated with the virus alone demonstrated significantly lower CD42b expression, ranging from 15-25% and 20-40% on day 11, respectively (**Fig 4.2.6a**). These findings indicate that viral transduction may negatively impact cellular differentiation potential, potentially causing toxic effects (**Fig 4.2.6a**). Additionally, viability and cell count were greatly diminished in the edited and virus-alone treated samples compared to the gRNA-alone treated sample (**Fig 4.2.7 a & b**), suggesting toxicity induced by transduction.

Consistent with previous HDR studies in HEL cells, the SA-P2A-GFP donor exhibited lower knock-in efficiency compared to the IRES-GFP donor in CD34+ cells (**Fig 4.2.6b**). HDR efficiency decreased over time, although CD41a expression increased, potentially due to toxicity associated with viral transduction (**Fig 4.2.6b**). Interestingly, HDR persisted in cells differentiated into mature megakaryocytes (CD41+

and CD42+ cells) (**Fig 4.2.6b**). To mitigate viral transduction toxicity, a temporal validation of the transduction procedure was performed using an AAV6-IRES-GFP donor with a fixed viral titre of 60K MOI. The study aimed to determine the optimal transduction time for achieving optimal knock-in efficiency while limiting viral exposure. After nucleofection of HEL cells with the RNP complex, cells were subjected to viral addition. Subsequently, the virus was eliminated by replacing the culture media with fresh media at different time intervals ranging from 2 to 48 hours. Knock-in efficiency was then evaluated using flow cytometry after 48 hours (**Fig 4.2.6c**). In the HEL cell line experiment, a plateau in HDR efficiency was observed after 12 hours of viral transduction, suggesting that removing the virus after 12 hours may reduce toxicity associated with megakaryocytic differentiation. Additionally, we tested whether HDR efficiency could be improved by comparing viruses obtained at 72 hours and 144 hours. We observed a significant increase in knock-in efficiency when using the 72-hour virus (**Fig 4.2.6d**). However, the drawback of using the 72-hour virus is the lower titre achieved compared to the 144-hour virus (**Fig 4.2.7 f&g**).

To validate the previous observation, CD34+ cells were nucleofected with the RNP complex and transduced with AAV collected at 72 hours. We used a 60K MOI IRES-GFP donor template for transduction and removed the virus after 12 hours. Interestingly, good differentiation was observed on day 10 of megakaryopoiesis. There was no significant difference between the unedited CD34+ cells and virus-alone treated samples, while gRNA-edited and knocked-in samples showed a reduction in the differentiation pattern, possibly due to indel creation and stress related to HDR-mediated repair pathway (**Fig 4.2.6e**).

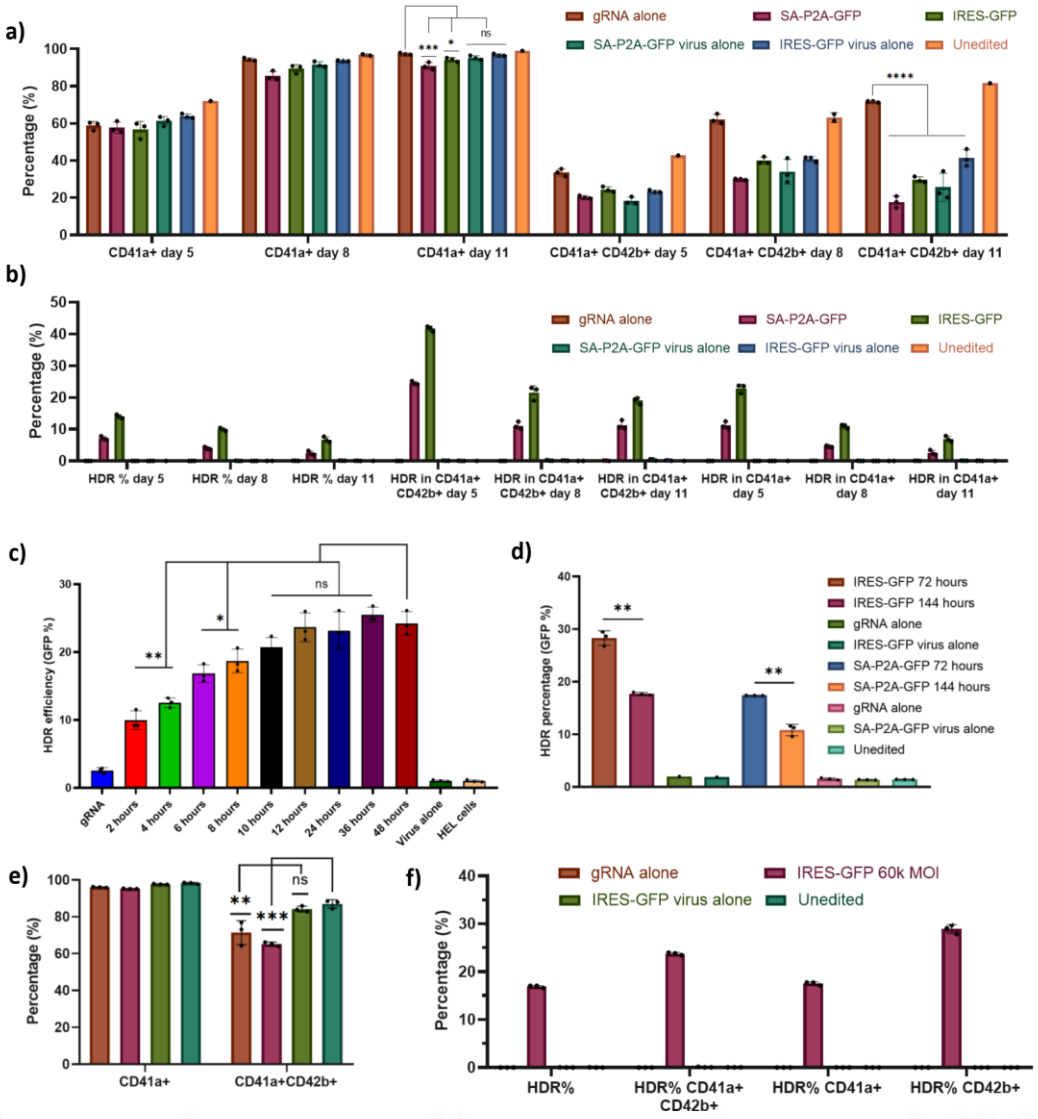


Figure 4.2.6: Enhancement of HDR-mediated gene knock-in efficiency under the ITGA2B promoter while minimizing CD34+ cell toxicity by manipulating delivery parameters: a) Evaluation of the megakaryocyte differentiation profile on days 5, 8, and 11 after HDR in CD34+ cells using a 30k MOI with SA-P2A-GFP and IRES-GFP donor constructs. b) Assessment of HDR efficiency in the overall population, megakaryocytes (CD41a), and matured megakaryocytes (CD41a+ CD42b+) after HDR in CD34+ cells using a 30k MOI with SA-P2A-GFP and IRES-GFP donor constructs. c) Time-dependent validation of HDR using the IRES-GFP construct at a 60k MOI in Hel cell lines: Identification of the optimal time point for virus removal post-transduction to mitigate toxicity while maintaining HDR efficiency. d) Comparative analysis of HDR efficiency between the SA-P2A-GFP and IRES-GFP donor constructs in the HEL cell line at a 30k MOI, using virus collected at 72 hours and 144 hours. e) Assessment of the megakaryocyte differentiation profile on day 9 after HDR in CD34+ cells using a 60k MOI with IRES-GFP donor constructs, with virus removal at 12 hours post-transduction. f) Determination of HDR efficiency in the overall population, megakaryocytes (CD41a), and matured megakaryocytes (CD41a+CD42b+) following HDR in CD34+ cells using a 60k MOI of IRES-GFP donor constructs, with virus removal at 12 hours post-transduction. The data are presented as mean ± SEM from three experimental replicates, and asterisks indicate levels of statistical significance: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

Although there was an observable enhancement in viability and cell count, a lesser degree of improvement was observed in the knocked-in sample and the gRNA-alone treated samples. These findings suggest that the observed toxicity is not solely attributed to viral transduction but also to the editing procedure itself (**Fig 4.2.7 a-d**). The overall HDR percentage obtained was above 15%, which was higher in mature megakaryocytes (<25%) since GFP was expressed under a megakaryocytic promoter (**Fig 4.2.6f**). However, the HDR efficiency was lower than that achieved in the HEL cell line (**Fig 4.2.3d**), which was around 35%. High indel efficiency was observed before knock-in in both HEL and CD34 cells (**Fig 4.2.7e**).

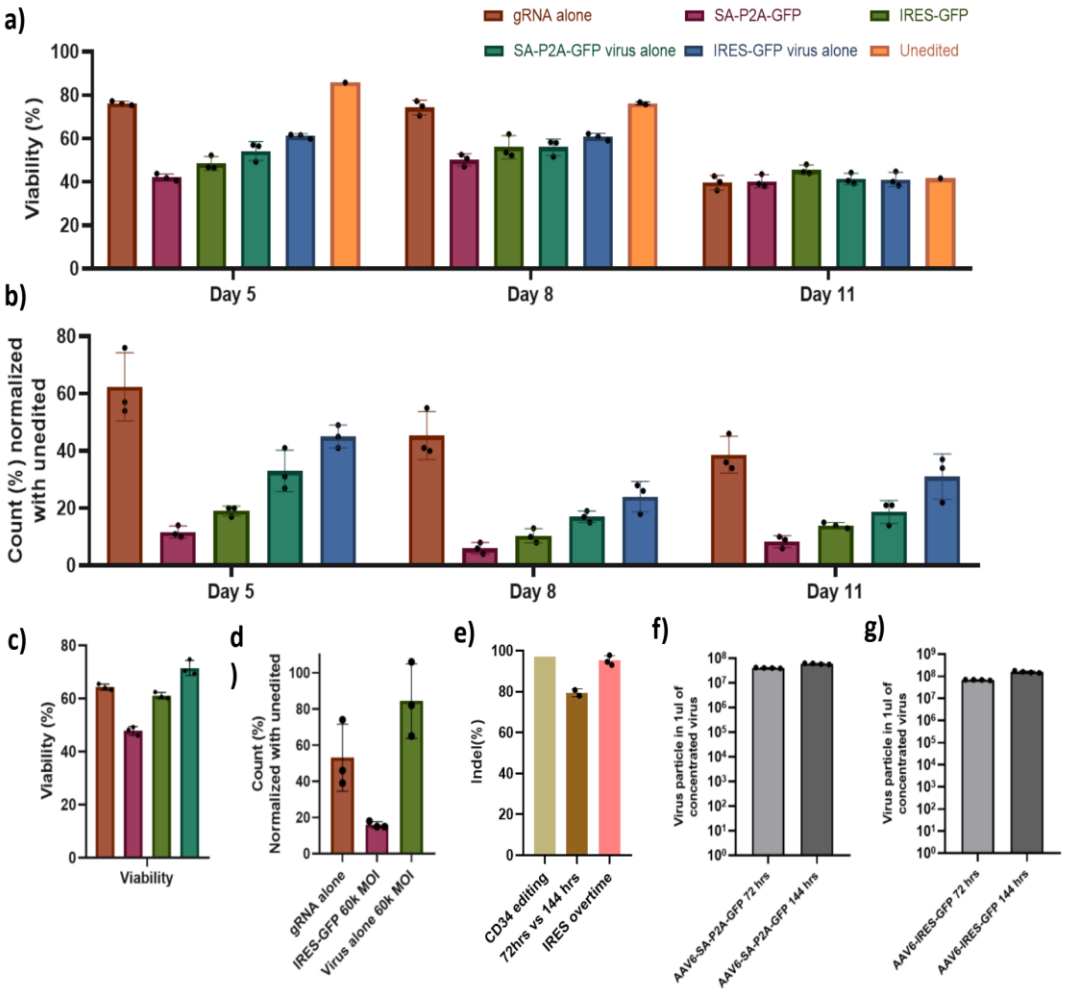


Figure 4.2.7: Assessment of cell viability and cell count in HDR-mediated gene editing using SA-P2A-GFP and IRES-GFP constructs in hematopoietic stem and progenitor cells (HSPCs). Evaluation of viability (a) and percentage of cell count (b) on days 5, 8, and 11 during megakaryocytic differentiation following HDR in CD34+ cells using a 30k MOI with SA-P2A-GFP and IRES-GFP donor constructs. Assessment of viability (c) and percentage of cell count (d) on day 9 during megakaryocytic differentiation following HDR in CD34+ cells using a 60k MOI with SA-P2A-GFP and IRES-GFP donor construct. e) Analysis of indel efficiency on day 4 after HDR editing using gRNA IT10. Quantitative evaluation of AAV6 containing the SA-P2A-GFP (f) and IRES-GFP (g) donor template through relative quantification (ΔΔCT) using quantitative reverse transcription PCR (qRT-PCR).

4.2.4 Insights into megakaryopoiesis and platelet biogenesis revealed by gene editing of the ITGA2B locus in HSPCs:

To enhance our understanding of megakaryopoiesis and platelet biogenesis, we conducted experiments involving gene editing of the ITGA2B locus in CD34+ cells. In order to reduce the toxicity associated with viral transduction, we modified our experimental approach. We increased the cell count to 0.5 million cells per milliliter to enhance cell density and potentially mitigate the stress induced by gene editing, specifically the toxicity related to electroporation and the HDR-mediated repair pathways. The objective was to assess whether higher cell density could effectively decrease toxicity. For this experiment, we utilized an MOI of 60K with the IRES-GFP donor template. **Figure 4.2.8a** illustrates the process of megakaryopoiesis and platelet biogenesis. Following transduction, we changed the media at the 12-hour mark and collected a small number of cells at 48 hours for quick DNA extraction. Sanger sequencing was performed to determine the indel efficiency. Satisfactory indel rates exceeding 90% were achieved in the gRNA alone and HDR-edited samples, while no editing was observed in the virus alone treated samples (**Fig 4.2.8b**). On day 6, we isolated DNA using a kit and assessed the knock-in efficiency using ddPCR, resulting in an integration rate of 30% (**Fig 4.2.8c**).

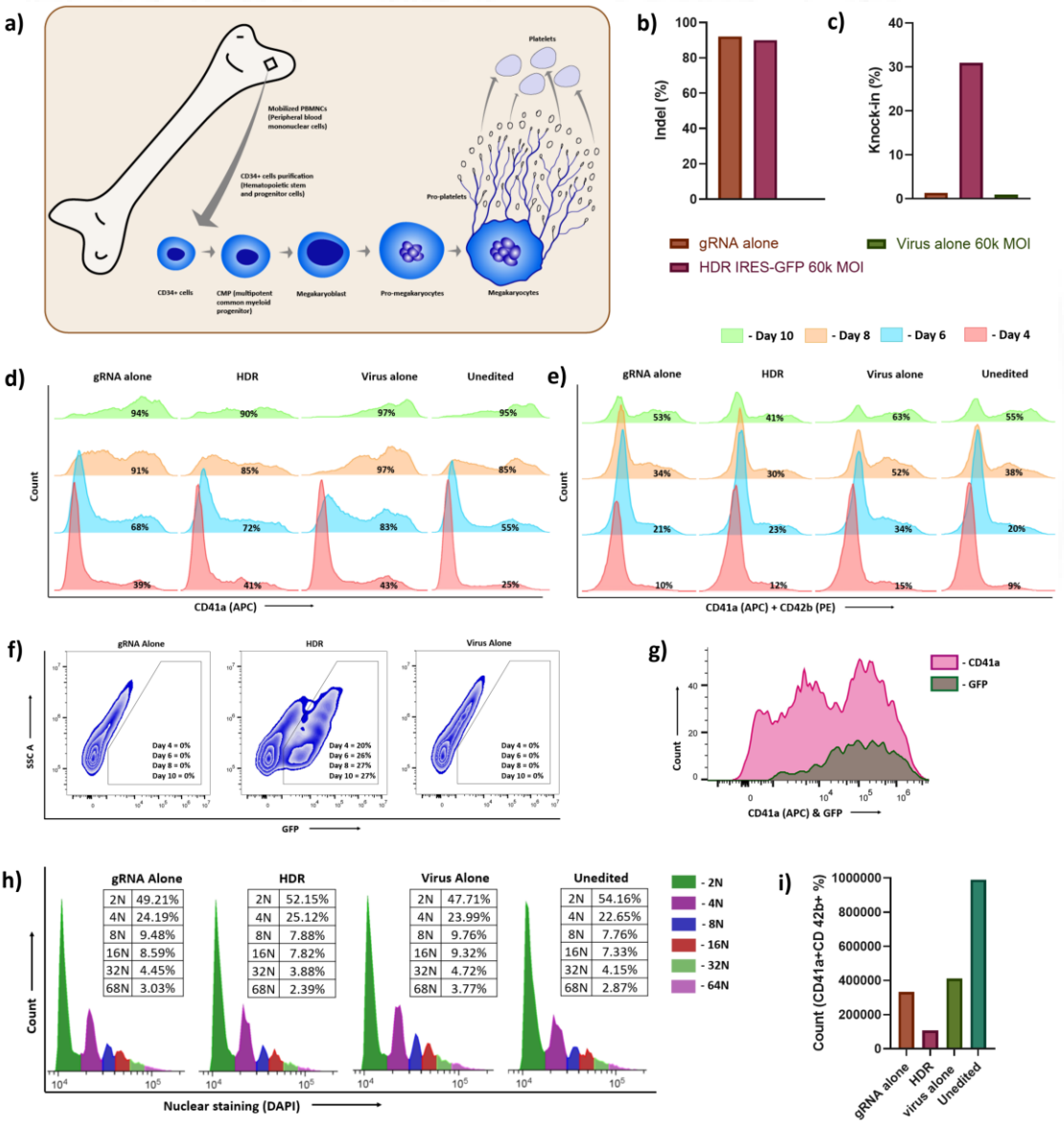
To validate megakaryopoiesis, measure viability, cell count, and assess knock-in efficiency based on GFP expression, we collected samples on days 4, 6, 8, and 10,

and analyzed them using flow cytometry. Flow cytometry analysis focused on examining populations expressing CD41a+ (an early megakaryocytic marker) and CD41a+CD42b+ (a late megakaryocytic marker). The percentage of CD41a-expressing cells increased from 40% on day 4 to 90% on day 8, indicating enhanced megakaryopoiesis in the edited sample compared to the unedited control (**Fig 4.2.8d and Fig 4.2.9e**). Notably, the virus alone treated sample showed over 80% CD41a expression on day 6, suggesting that the virus may promote megakaryopoiesis. The MFI of CD41a+ cells also increased over time, with the virus alone treated sample exhibiting higher MFI (**Fig 4.2.9c**). Regarding CD41a+CD42b+ expression, above 50% expression was observed on day 10. Similar to CD41a expression, the virus alone treated sample exhibited faster megakaryocytic maturation compared to other samples and the unedited control (**Fig 4.2.8e**). The expression of GFP, which was knocked-in at the ITGA2B locus, increased from 20% on day 4 to 27% on day 10. A notable rise from 20% on day 4 to 26% on day 6 suggests a correlation between the increase in CD41a+ expressing cells and enhanced GFP expression. Importantly, there was no decrease in the percentage of GFP-positive cells, indicating a reduction in toxicity associated with viral transduction (**Fig 4.2.8f**). In terms of viability, all samples exhibited lower viability compared to the unedited cells, with viability rates above 50%. Among the edited samples, the knocked-in sample displayed the lowest viability (**Fig 4.2.9a**). When considering the cell count percentage, a significant decrease was observed in the edited sample compared to the unedited sample (**Fig 4.2.9b**). Despite increasing the cellular density from 0.1 to 0.5 million cells per ml, the enhancement in cell count or viability was not substantial. Among the edited samples, the virus alone treatment showed a higher cell count percentage of approximately 50% compared to the unedited sample.

The gRNA-treated samples ranged between 25-40% across different days, while the knocked-in sample exhibited less than 20% cell count percentage (**Fig 4.2.9b**).

Although cell viability appears satisfactory, the data regarding cell count suggests significant toxicity resulting from the gene editing process. To determine whether the observed decrease in cell numbers resulted directly from editing (specifically, the replacement in the ITGA2B locus) or from broader processes such as electroporation, transduction, and the HDR pathway, we conducted a series of analyses. We assessed the mean fluorescence intensity (MFI) of the entire CD41a population to evaluate the potential impact on per cell CD41a (ITGA2B) expression in the edited cells. As expected, the MFIs of the virus alone and gRNA alone samples were higher than those of the unedited sample, indicating that editing and viral treatment stimulated megakaryopoiesis. While the knocked-in sample displayed a lower MFI compared to other edited samples, it was comparable to the unedited sample, suggesting that editing does not affect per cell CD41a expression (**Fig 4.2.9c**). To further validate the CD41a MFI in the CD41a+GFP+ and CD41a+GFP- populations of the knocked-in sample, we conducted an analysis to assess whether GFP-expressing cells exhibit reduced CD41a expression. Surprisingly, the CD41a+GFP+ population exhibited significantly higher expression compared to the CD41a+GFP- population (**Fig 4.2.9d**), further supporting the notion that the editing process does not negatively impact the cells. On day 10, CD41a expression in the knocked-in samples exhibited a distinct peak with low CD41a expression and another peak with high CD41a expression (**Fig 4.2.8d**). To verify that the low-intensity peak was not influenced by the knock-in effect, we conducted an overlay analysis of the CD41a+ histogram with the GFP+ histogram. Surprisingly, the GFP events overlapped with the high peak of CD41a+ rather than the low peak,

indicating that the low expression peak of CD41a is not a consequence of the editing process (**Fig 4.2.8g**). Additionally, this finding showcases the synchronized expression of GFP and the ITGA2B gene within the CD41a+ GFP+ population, which remains stable without a decrease in frequency over time (**Fig 4.2.9e**). In conclusion, the decrease in cell count may not solely be attributed to the editing itself but rather to associated procedures such as electroporation (Batista Napotnik *et al*., 2021) followed by viral transduction and the stress associated with the HDR pathway, which could potentially be mitigated by the application of appropriate apoptotic inhibitors.



gRNA Alone	
2N	49.21%
4N	24.19%
8N	9.48%
16N	8.59%
32N	4.45%
68N	3.03%

HDR	
2N	52.15%
4N	25.12%
8N	7.88%
16N	7.82%
32N	3.88%
68N	2.39%

Virus Alone	
2N	47.71%
4N	23.99%
8N	9.76%
16N	9.32%
32N	4.72%
68N	3.77%

Unedited	
2N	54.16%
4N	22.65%
8N	7.76%
16N	7.33%
32N	4.15%
68N	2.87%

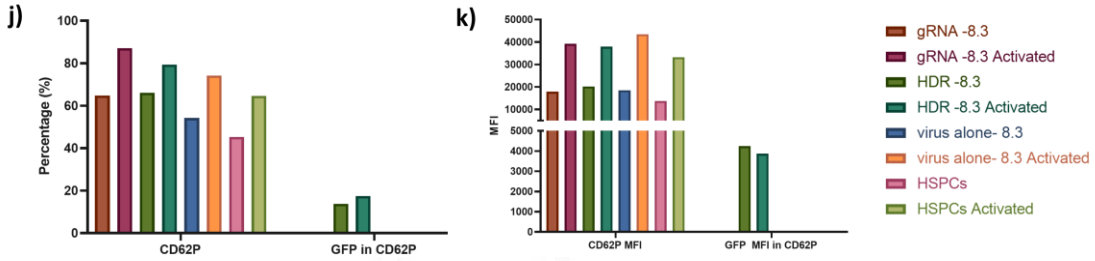


Figure 4.2.8: Insights into megakaryopoiesis and platelet biogenesis derived from gene editing at the ITGA2B locus in hematopoietic stem and progenitor cells (HSPCs): Verification of knock-in efficiency and evaluation of platelet functionality. a) Diagram illustrating the process of megakaryopoiesis. CD34+ cells were subjected to gene editing with an AAV6-IRES-GFP donor template at a 60k MOI for HDR. b) Calculation of the percentage of indels obtained from editing the ITGA2B locus using gRNA IT10. c) Confirmation of knock-in efficiency through droplet digital PCR (ddPCR) in CD34+ cells edited with the IRES-GFP donor template at the ITGA2B locus. Percentage of megakaryocytes (CD41a+) (d) and mature megakaryocytes (CD41a+CD42b+) (e) generated during megakaryocytic differentiation on days 4, 6, 8, and 10. f) Verification of knock-in efficiency by flow cytometry through GFP expression in CD34+ cells edited with the IRES-GFP donor template at the ITGA2B locus. g) Overlayed histogram displaying CD41a+ cells co-expressing GFP obtained from the HDR sample. h) Investigation of heterogeneity in megakaryocytic ploidy: Quantitative analysis of chromosomal numbers on day 10 of differentiation. i) Enumeration of platelets by counting CD41a+ CD42b+ populations in gRNA alone, HDR, virus alone, and unedited samples on day 10 of megakaryopoiesis using flow cytometry. j) Quantitative analysis of platelet activation by CD62p immunostaining and co-expression of GFP in response to TRAP stimulation. k) Estimation of the mean fluorescence intensity (MFI) of CD62p and GFP expression before and after platelet activation

To validate the production of mature megakaryocytes during megakaryopoiesis, immune phenotyping alone is insufficient. Therefore, we conducted ploidy analysis, as ploidy serves as a characteristic feature of megakaryocytes. Ploidy analysis was performed on the CD41a+high & GFP+ population of edited samples on day 10, revealing ploidy levels up to 64N in all samples. These ploidy levels were comparable to those observed in the unedited sample (**Fig 4.2.8h**). Furthermore, the ploidy of the CD41a+high GFP- and CD41a+high GFP+ populations within the knocked-in sample was analyzed to ensure that editing did not impact megakaryocyte maturation. Surprisingly, the CD41a+high GFP+ population displayed favorable ploidy levels compared to the GFP- population, further confirming the absence of any adverse events impeding megakaryopoiesis following the knock-in of the gene of interest (GOI) at the ITGA2B locus (**Fig 4.2.9f**). Following the confirmation of megakaryocyte maturation,

proplatelet formation was examined. Proplatelets are elongated cytoplasmic extensions produced by mature megakaryocytes, leading to platelet production. On day 11, proplatelets were observed in all samples (**Fig 4.2.9g**). In concurrence with a previous study, our observations during megakaryopoiesis enabled us to discern the presence of ploidy variations and the formation of proplatelets (Perdomo *et al*., 2017). However, the edited samples exhibited a lower number of proplatelets compared to the unedited samples, possibly due to reduced cell count. Additionally, proplatelets in the unedited samples appeared longer than those in the edited samples.

The presence of proplatelets indicates the forthcoming production of platelets. However, distinguishing mature megakaryocytes from platelets is challenging due to their shared expression of CD41a and CD42a/b markers. Size and granularity are the primary distinguishing factors, as platelets are smaller and exhibit reduced granularity compared to megakaryocytes. However, it is important to note that platelets generated *In Vitro* during megakaryopoiesis may not precisely resemble those produced in vivo in terms of shape and size. To enumerate platelet production under *In Vitro* conditions, we considered the total population of CD41a+CD42b+ cells. As anticipated, a significant decrease in the double-positive population was observed in the edited samples compared to the unedited control (**Fig 4.2.8i**). This decrease can be attributed to the overall lower count of edited cells compared to the unedited sample (**Fig 4.2.9b**). Moreover, the double-positive population was substantially lower in the knocked-in sample, although this cannot be directly attributed to editing, as the percentage of CD41a+CD42b+ cells on day 10 did not vary much compared to the unedited sample (**Fig 4.2.8e**).

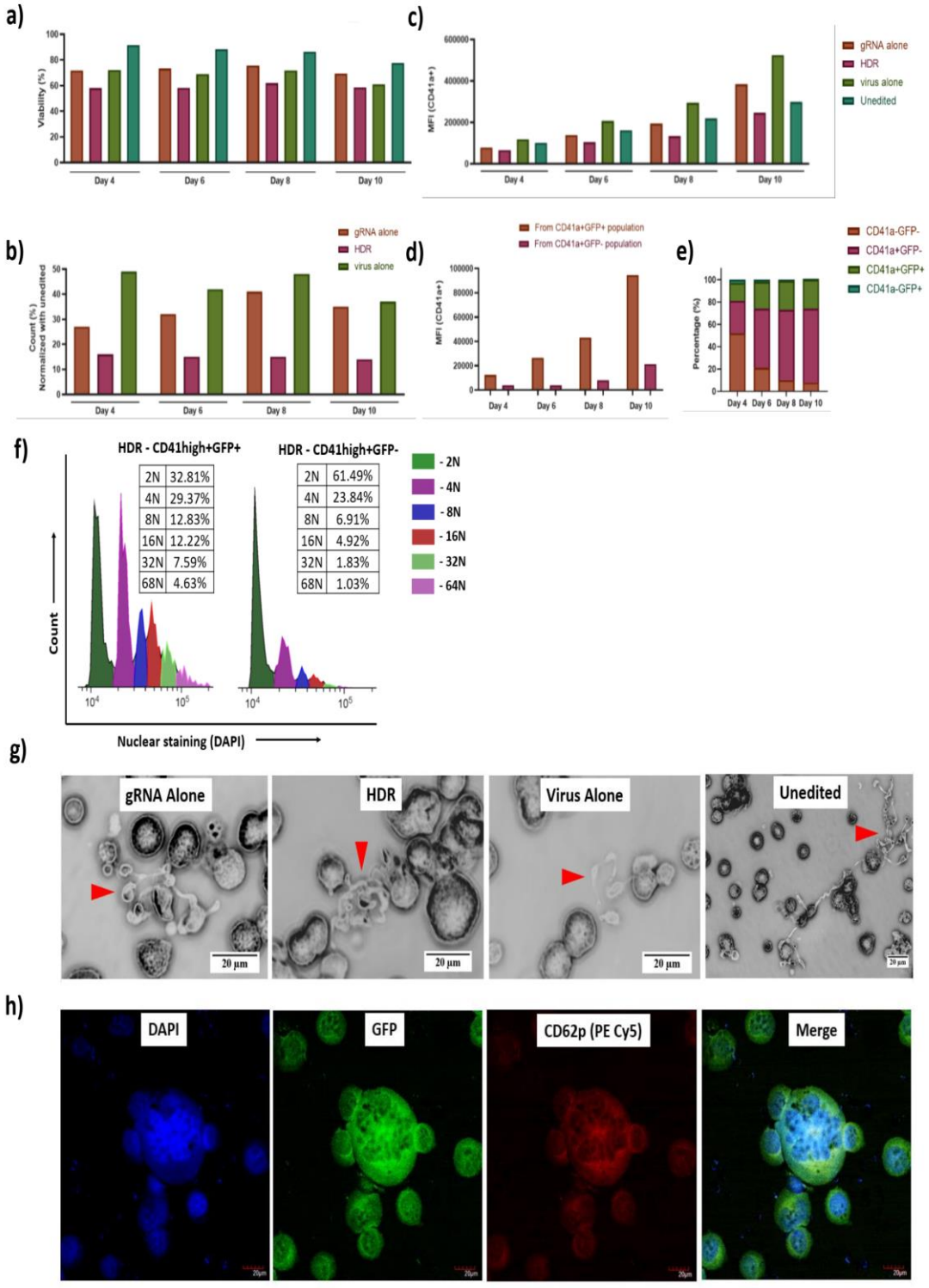


Figure 4.2.9: Characterization of megakaryopoiesis in edited CD34 cells: Evaluation of viability, cell *count*, and proplatelet formation, with implications for GFP co-localization in alpha granules of the HEL cell line. Assessment of viability (a) and percentage of cell count (b) on days 4, 6, 8, and 10 during megakaryocytic differentiation following HDR in CD34+ cells using a 60k MOI IRES-GFP donor construct. c) Mean fluorescence intensity (MFI) of CD41a+ expression in the edited sample throughout megakaryopoiesis. d) MFI of CD41a+ cells co-expressing GFP in the HDR sample. e) Identification of different population types observed during megakaryopoiesis. f) Ploidy analysis in GFP+ and GFP- cells

from the HDR sample. g) Observation of proplatelets through bright-field microscopy on day 11 of megakaryopoiesis in all experimental samples. h) Qualitative analysis of co-localization between knocked-in GFP at the ITGA2B locus and CD62p in alpha granules of megakaryoblast cells obtained from the edited HEL cell line using confocal microscopy.

Despite the inability to discern platelets from mature megakaryocytes based on immunophenotyping, we employed a gating strategy to select a specific population for platelet activation study. This population was determined by their smaller size and lower granularity resembling platelets, as indicated by low FSC and SSC. On day 13 of megakaryopoiesis, platelet functional analysis was conducted through a platelet activation assay. Following gating of the CD41a+CD42a+ population, activated platelets were assessed by staining with CD62p antibody after treatment with TRAP (thrombin receptor activating peptide). Interestingly, CD62p staining was observed in both inactivated and activated samples, as platelets derived from *In Vitro* megakaryopoiesis undergo significant stress. Nonetheless, a clear increase in CD62p expression was evident after activation (**Fig 4.2.8j**). Notably, GFP expression was observed within the CD62p population of the HDR sample, confirming the presence of GFP within platelets. Furthermore, the MFI for CD62p drastically increased following activation. Intriguingly, in HDR samples, the GFP MFI slightly decreased after activation, suggesting the release of stored proteins from α-granules during activation (**Fig 4.2.8k**). To validate the localization of GFP within α-granules, Hel cells with AAV6-IRES-GFP knock-in were permeabilized and stained with CD62p. Given that CD62p is known to be localized in α-granules, a distinct colocalization between CD62p (PEcy5) and GFP was observed in the knocked-in Hel cells (**Fig 4.2.9h**). In summary, the decrease in cell viability and count observed in edited samples can be attributed to the stress induced by the gene editing process and associated procedures, rather than the direct effect of editing. Furthermore, the successful generation of mature

megakaryocytes, proplatelets, and functional platelets demonstrates the potential of gene editing in enhancing megakaryopoiesis and platelet biogenesis.

4.2.5 Coordinated Expression of Factor IX and ITGA2B through Targeted Gene Editing:

Following the successful establishment of coordinated expression between GFP and ITGA2B without compromising megakaryopoiesis, we proceeded to replace the GFP coding sequence with the FIX PADUA coding sequence through targeted gene editing. This led to the creation of an adeno-associated virus serotype 6 (AAV6) vector carrying an internal ribosome entry site (IRES) and the FIX PADUA donor construct. **Figure 4.2.10a** provides a schematic representation of the monoallelic integration of FIX in the ITGA2B locus and details the structural composition of the AAV6-IRES-FIX donor construct. Validation experiments conducted in HEL cells demonstrated a high efficiency of indel formation (>90%) (**Fig 4.2.10b**) and a knock-in efficiency of 20% (**Fig 4.2.10c**) as confirmed by Sanger sequencing and droplet digital PCR (ddPCR), respectively. Analysis of relative RNA expression revealed a significant increase in FIX expression in the knocked-in samples compared to samples treated with gRNA alone or virus alone (**Fig 4.2.5d**). Notably, no substantial reduction in ITGA2B expression was observed at the RNA level. Encouraged by these results in HEL cells, we proceeded to employ the AAV6-IRES-FIX vector for FIX knock-in in CD34+ hematopoietic stem and progenitor cells (HSPCs) derived from healthy donors. The knock-in efficiency exceeded 30%, accompanied by an indel efficiency of over 90% (**Fig 4.2.10 e&f**). As expected, the edited cells exhibited lower viability and cell count percentages compared to the unedited sample (**Fig 4.2.11 a&b**).

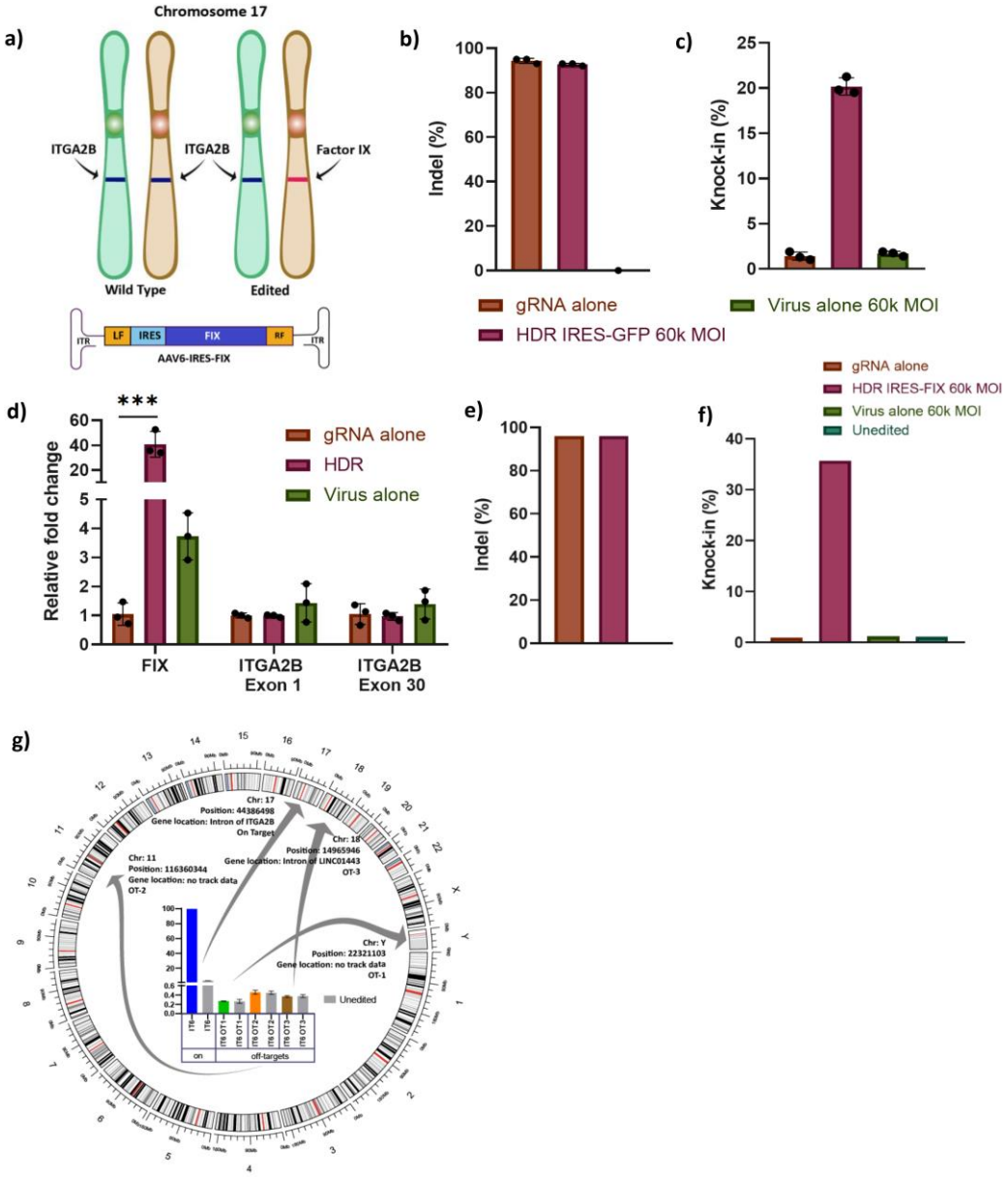


Figure 4.2.10: Validation and characterization of concerted expression of Factor IX (FIX) and ITGA2B through targeted gene editing: a) Illustration depicting the monoallelic insertion of FIX coding sequence (CDS) in the ITGA2B locus using the AAV6-IRES-FIX donor construct. b) Quantification of indel percentages obtained by editing the ITGA2B locus using gRNA IT10 in HEL cell lines. c) Validation of knock-in efficiency by droplet digital PCR (ddPCR) in HEL cell lines edited with the IRES-FIX donor template at the ITGA2B locus. d) Relative expression analysis of FIX and ITGA2B in HEL cells edited with the IRES-FIX donor template using quantitative real-time PCR (qRT-PCR). e) Determination of indel percentages obtained by editing the ITGA2B locus using gRNA IT10 in CD34+ cells. f) Validation of knock-in efficiency by ddPCR in CD34+ cells edited with the IRES-FIX donor template at the ITGA2B locus with a multiplicity of infection (MOI) of 60k. g) Analysis of Cas-dependent DNA off-target effects for the gRNA targeting the ITGA2B locus.

Upon differentiation of the edited cells into megakaryocytes, favorable megakaryopoiesis was observed, with over 80% of cells expressing CD41a (**Fig 4.2.11 c**). On day 10 of differentiation, the CD41a+CD42b+ population were quantified (**Fig 4.2.11 d**). Due to the low cell count of the edited cells and the inability to separate platelets from the culture system, human FIX ELISA or FIX chromogenic assay could not be conducted. Further investigations are necessary to transfer the edited cells into an ex vivo humanized mice model in order to obtain human platelets derived from the edited cells and accurately measure the levels of FIX. To evaluate the safety and specificity of the gRNA targeting the intron-1 of the ITGA2B locus, we conducted a comprehensive Cas-dependent DNA off-target analysis. Deep sequencing analysis was performed to identify potential off-target effects, but no significant off-target sites were detected at the three predicted sites (**Fig 4.2.10g**). Overall, the strategy of Monoallelic Integration with Cas9-mediated Endogenous (MICE) gene replacement holds promise for integrating a gene of interest (GOI) into other gene loci while preserving the expression of host genes. In this study, we successfully achieved monoallelic integration and coordinated expression of GFP in the ITGA2B locus using AAV6-based donor constructs in CD34+ cells. This resulted in high GFP expression without impacting ITGA2B expression. Although the study did not directly demonstrate the selective delivery of engineered platelets encapsulating FIX to injury sites upon activation, previous research has supported this aspect as illustrated in **Figure 4.2.11 e**. Additionally, the edited cells successfully differentiated into megakaryocytes and generated platelets expressing detectable levels of GFP without any differentiation defects.

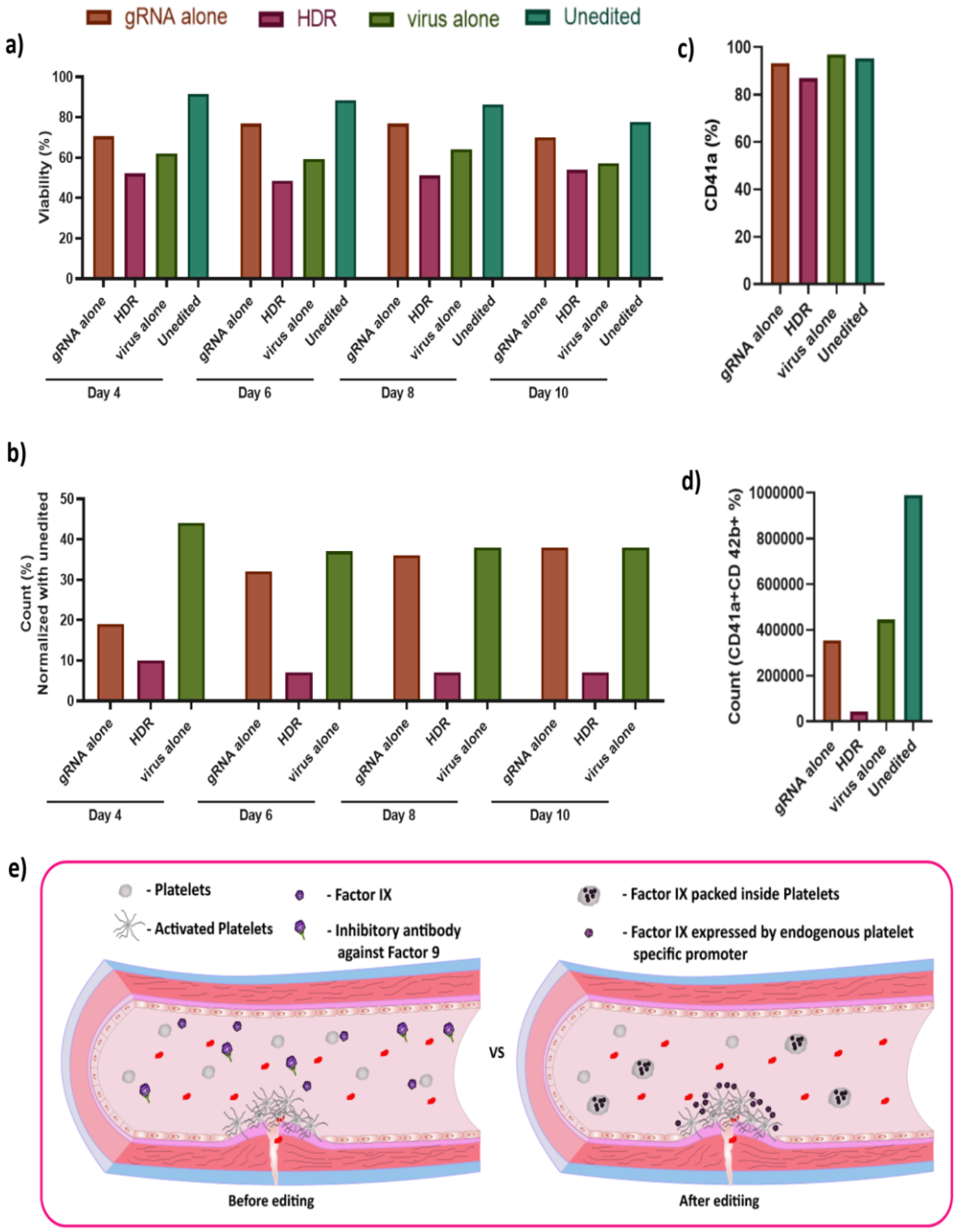


Figure 4.2.11: Characterization of gene replacement in CD34+ cells using engineered platelets carrying Factor IX (FIX) and enumeration of CD41a+CD42b+ cells during megakaryopoiesis: Viability percentage (a) and cell count percentage (b) assessed from days 4 to 10 following the editing of CD34+ cells using the AAV6-IRES-FIX donor template. c) Percentage of CD41a expression on day 10 post-editing. d) Enumeration of CD41a+CD42b+ edited cells on day 10 of megakaryopoiesis, prior to the cells being taken for FIX chromogenic assay. e) Schematic illustration depicting the presence of engineered platelets carrying FIX in circulation after implementing the novel gene replacement strategy known as MICE (Monoallelic Integration and Concerted Expression).

DISCUSSION

Section 5.1: Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin.

During the normal process of globin switching, the interaction between cis-acting elements and various transcription factors leads to the suppression of fetal globin and the activation of beta-globin (Ikuta *et al*., 1996). In order to gain a better understanding of how the gamma-globin gene is regulated, we employed two different base editing techniques to examine the HBG promoter at the level of individual nucleotides. Through this approach, we identified several new nucleotide substitutions in the HBG promoter that enhance the production of HbF by modifying the binding sites for transcriptional activators or repressors.

Existing methods for studying the regulation of fetal globin using programmable nucleases often result in the deletion of the HBG2 gene by introducing double-strand breaks in both HBG promoters (Traxler *et al*., 2016). This removal of the 4.9 kb intergenic region, which includes the HBG2 gene, allows the locus control region (LCR) to directly interact with the HBG1 promoter and activate its expression (M′etais *et al*., 2019). However, it can be challenging to determine the specific impact of different mutations associated with hereditary persistence of fetal hemoglobin (HPFH) on gamma-globin expression, as these mutations can occur in either or both HBG2 and HBG1 promoters. Additionally, the use of CRISPR-Cas9-based editing often leads to a variety of insertions and deletions at the target sites, making it difficult to pinpoint the exact mutations involved in gene regulation. In contrast, our base editing strategy allows us to modify target bases within a specific editing window without creating double-

strand breaks, thereby minimizing disruptions to the HBG locus. By employing this strategy, we successfully targeted regions in both HBG1 and HBG2 promoters and efficiently edited the sites with fewer or no significant deletions, enabling us to evaluate gamma-globin expression from both active promoters.

Utilizing base editors incorporating a nickase variant of the CRISPR/Cas9 system, we observed a minor proportion of 4.9 kb deletions. The occurrence of these larger deletions is primarily attributed to the simultaneous induction of double-stranded DNA breaks by the CRISPR-Cas9 system or the generation of two single-strand breaks (SSBs) on opposite DNA strands of the HBG1 and HBG2 gene through paired nickase activity (Ann Ran *et al*, 2013). Intriguingly, recent investigations have unveiled the possibility of adjacent SSBs on the same DNA strand leading to the formation of genomic deletions in plant species. The frequency of deletions is contingent upon the initial liberation of the single-stranded fragment located between the two SSBs (Schiml *et al*., 2016). Moreover, additional reports propose the conversion of persistent nicks into double-strand breaks facilitated by the replication fork. When the replication fork, primed by an R-loop, encounters the single-strand nick site on the DNA template, it undergoes a collapse event that culminates in the production of a double-strand break (Kuzminov, 2001; Wimberly *et al*., 2013). Based on these findings, we hypothesize that the simultaneous introduction of SSBs by base editors at the editing site of the HBG1 and HBG2 promoter may engender a subset of larger 4.9 kb deletions, although we observed minimal instances of such occurrences.

Numerous HPFH point mutations have been documented within the HBG promoters, and their impact on gamma-globin expression in the native cellular context has been elucidated for a limited subset of these mutations (Daniel E. Bauer and Stuart H.Orkin, 2012; Liu *et al*., 2018; Wienert *et al*., 2015, 2017). Our findings align with

previous studies demonstrating that point mutations in three distinct regions of the HBG promoters, specifically positions -198, -175, and -115, mimic HPFH-associated mutations that affect crucial regulators of HbF expression (Liu *et al*., 2018; Martyn *et al*., 2018; Stoming TA *et al*, 1989; Wienert *et al*., 2015, 2017). Among the known HPFH mutations, base conversion within the -115 cluster (ranging from -110 to -116) exhibited the most pronounced increase in promoter activity, corroborating earlier investigations (Fucharoen *et al*., 1990; J.G.Gilman *et al*, 1988; Motum *et al*, 1994; S. Zertal-Zidani, Merghoub *et al*., 1999). Using cytidine base editors (CBEs), we observed that base conversion (C to T) at positions -114 and -115 resulted in significantly higher induction of HbF compared to multiple A to G nucleotide substitutions at positions -110, -112, -113, and -116 performed using adenine base editors (ABEs). Recent studies have demonstrated that the key HbF repressor, BCL11A, directly binds to the core TGACC motif located between positions -114 and -118 (Liu *et al*., 2018; Martyn *et al*., 2018). Naturally occurring HPFH mutations at -117G>A, -114C>A, -114C>T, -114C>G, and a 13bp deletion (Δ13bp) disrupt the binding of BCL11A to the promoter (Martyn *et al*., 2018). The -113A>G HPFH mutation within the -115 cluster creates a binding site for the master erythroid regulator GATA1 without disrupting BCL11A binding (Martyn *et al*., 2019). Our results align with previous reports, indicating that perturbation of the core binding region of BCL11A and the creation of novel binding sites for GATA1 result in increased fetal globin expression in wild-type HUDEP-2 cells (Wienert *et al*., 2017). ABE-mediated T to C substitution at position -198 of the HBG gene promoter has previously been associated with British HPFH and significantly enhances HbF expression by generating a de novo binding site for the erythroid gene activator KLF1 (VE Tate, 1986; Wienert *et al*., 2017). Another well-known HPFH mutation (-175T>C) has been shown to facilitate enhancer looping to the HBG promoter through recruitment

of the activator TAL1 (Wienert *et al*., 2015). Furthermore, improved editing efficiency at the -175 T>C position using hyperactive variants of ABE (ABE 8e) resulted in the highest induction of HbF synthesis in human erythroid cells.

In this investigation, we have discovered several novel point mutations in the HBG promoter associated with elevated HbF levels. Through ABE-mediated base editing, we identified multiple potential regions involved in HbF regulation in the HBG promoter, as ABE exhibited greater compatibility with the targeted region compared to CBE due to the availability of more ABE-compatible gRNAs. Alongside the known mutations, we identified substitutions at positions -69 (C to T), -70 (C to T), -122 (G to A), -123 (T to C), and -124 (T to C) of the HBG promoters as potential new regulatory mutations capable of enhancing gamma-globin expression. The observed levels of gamma-globin expression resulting from these mutations closely resembled those of well-characterized, naturally occurring HPFH mutations. Our study predicted that nucleotide substitutions at positions -123 T>C and -124 T>C within the HBG promoter might reactivate gamma-globin expression by creating a binding site for KLF-1, which was subsequently confirmed through EMSA. This finding, in conjunction with the observation that the -198 T to C mutation forms a de novo KLF1 site capable of upregulating fetal globin (VE Tate, 1986; Wienert *et al*., 2017), suggests that the introduction of a KLF1 binding site anywhere in the vicinity of the HBG promoter could potentially enhance HBG gene expression. However, in contrast to our EMSA findings, the ChIP analysis only revealed a weak signal for KLF1 binding at the edited site of the HBG promoter. Therefore, our hypothesis, primarily based on *In Vitro* binding of KLF1 in EMSAs, proposes that the -123 and -124 mutations create a new, relatively weak KLF1 binding site that is challenging to detect using ChIP. However, alternative hypotheses are also plausible, such as the creation of a binding site for another activator. Given the close proximity of this site to the BCL11A binding site, which begins around

-117, it is possible that it may directly or indirectly influence BCL11A binding. Further investigations are necessary to explore these possibilities.

The screening methods employed in our study to identify the regulatory element in the proximal promoter of the HBG promoter have certain technical limitations. The presence of NGG PAM sequences in the target region restricts the resolution of the screening approach. Furthermore, the editing efficiency of ABE7.10 RA or BE3RA-FNLS is not uniform across the target regions, as highlighted in previous studies (Koblan *et al*., 2018). Additionally, the current base editors primarily facilitate transition mutations, thereby making it impossible to assess the impact of transversion mutations on gene regulation (Nicole M Gaudelli *et al*., 2017; Komor, Kim, Packer, Zuris and Liu, 2016). The introduction of bystander mutations by base editors at the target regions further complicates the identification of functional regulatory single-nucleotides responsible for gamma-globin regulation. To address these limitations, alternative strategies can be employed. This includes the utilization of base editor variants that recognize non-canonical PAM sites, as described by Richter *et al*. Moreover, the use of recently developed hyperactive base editor variants can enhance editing efficiency at the target site with a broader editing window (Richter *et al*., 2020). The scope of our study can be expanded by incorporating the dual adenine and cytosine base editor, which enables simultaneous conversions of A-G and C-T, as well as by employing prime editing techniques that offer a wider range of precise conversions within the desired region (Anzalone *et al*., 2019; Zhang *et al*., 2020).

The clinical applicability of genetically modified hematopoietic stem and progenitor cells relies on their ability to engraft and repopulate long-term. However, the occurrence of genotoxicity and cytotoxicity resulting from double-strand breaks (DSBs) induced by programmable nucleases can pose limitations (Cullot *et al*., 2019; Kyung-Rok Yu *et al*, 2016). A previous study conducted in nonhuman primates observed that

Cas9-mediated editing of the HBG promoter led to HBG2 deletions with a frequency of up to 27%, and cells with this deletion were found to be under-represented after engraftment (Humbert *et al*., 2019). On the other hand, when targeting the HBG1 and HBG2 promoter sites using adenine base editors (ABEs) and cytosine base editors (CBEs), there was no significant occurrence of large deletions in the intergenic region, as observed with Cas9. Furthermore, only low levels of indel formation were observed. ABEs possess an inherent advantage over CBEs as they exhibit high fidelity in generating the desired edits (A:T to G:C), while CBEs can result in unanticipated edits. Consistent with previous findings, our results also support the notion that ABE outperforms CBE in terms of achieving accurate base conversion and minimizing indel formation (Lee *et al*., 2018). Moreover, our preliminary results indicate that base editing of hematopoietic stem and progenitor cells using the ABE 8e variant, targeting a novel site with gRNA 11, significantly elevated HbF levels in erythroid progeny, potentially reaching therapeutic levels. Importantly, no significant off-target effects in terms of DNA or RNA alterations were observed in cells edited using ABE or CBE. Our proof-of-principle study validates the effectiveness of various gRNAs in elevating HbF levels to therapeutic ranges, laying the foundation for potential clinical applications. This approach holds promise for addressing a range of β-globin disorders, obviating the need to develop specific therapeutic products for each individual disorder.

In conclusion, we have successfully demonstrated the efficacy of CRISPR base editing in inducing the expression of fetal hemoglobin at therapeutically relevant levels in both an erythroid progenitor cell line and hematopoietic stem and progenitor cells (HSPCs). Through an extensive screening process encompassing all gRNAs within the 320 bp region of the HBG promoter, we have identified nine gRNAs that, when used in conjunction with the appropriate base editor, can introduce HPFH-like mutations without causing deletions or insertions. Furthermore, our study has uncovered five

previously unidentified regulatory regions in HBG1 and HBG2 that play crucial roles in silencing gamma-globin expression in adult erythroid cells, thus shedding light on the underlying molecular mechanisms governing hemoglobin switching (see Figure 7). This work serves as an exemplification of the utility of base editors in elucidating gene regulatory elements within highly homologous loci. We anticipate that base editing strategies will emerge as one of the prominent therapeutic approaches for treating monogenetic disorders, including β-hemoglobinopathies, in the future.

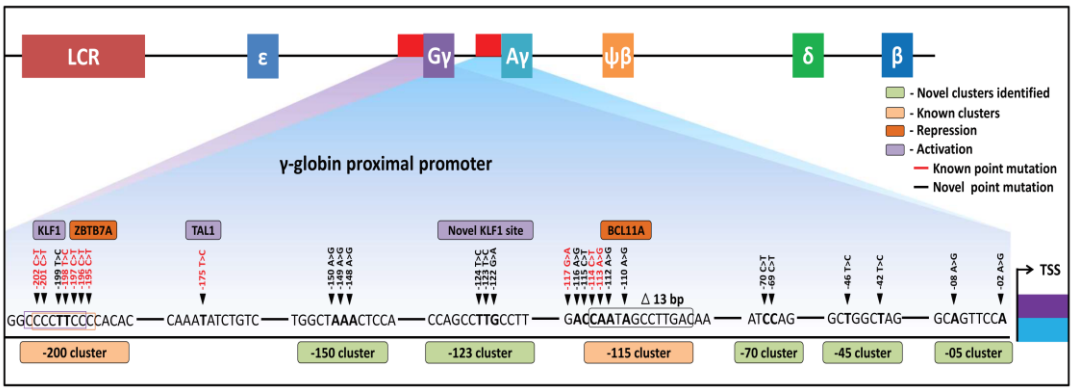


Figure 5.1.1 presents a schematic representation of the HBG promoter region, specifically spanning from the transcription start site (TSS) to -205 bases, focusing on the point mutations that contribute to the elevation of fetal hemoglobin (HbF). The proximal promoter regions of HBG2 and HBG1 are included in this depiction. In this study, we have identified novel clusters of mutations, highlighted in Sage color (five clusters), along with the previously known clusters highlighted in Melon color (two clusters). Known base conversions are depicted in black text, while the identified hereditary persistence of fetal hemoglobin (HPFH)-like mutations are indicated in red text. Furthermore, the novel base conversions discovered in our study are highlighted in bold font for emphasis. The figure also includes representations of transcriptional activators (lavender) and repressors (orange) that are known to bind to the identified clusters, providing additional insights into the regulatory mechanisms governing HbF expression.

Section 5.2: Monoallelic integration of Factor IX transgene in ITGA2B gene and their concerted expression can ameliorate haemophilia B

Gene replacement strategies have shown promising therapeutic potential for addressing genetic disorders (Azhagiri *et al*., 2021). However, conventional approaches that rely on the endogenous gene locus face limitations when the promoter region is mutated. To overcome these challenges, we have developed a novel gene replacement strategy called Monoallelic Integration and Concerted Expression (MICE) in this study. The MICE strategy aims to achieve synchronized gene expression within a shared biological process or pathway by precisely inserting a corrected gene sequence into a related gene locus associated with a similar biological pathway while preserving the expression of the host gene.

In order to evaluate the effectiveness of MICE, we employed the ITGA2B promoter to drive selective expression of coagulation factors (FIX) specifically within platelets. Our main objective was to achieve coordinated expression of both coagulation FIX and ITGA2B, as they play crucial roles in proper blood clotting and platelet function. Through precise targeting of the ITGA2B gene locus in hematopoietic stem and progenitor cells (HSPCs), we successfully integrated the corrected FIX coding sequence. By leveraging the regulatory potential of the ITGA2B promoter, FIX expression was exclusively induced within the lineage of megakaryocytes. Consequently, we generated engineered platelets containing FIX, which can be delivered to specific sites through platelet activation, ensuring the maintenance of hemostasis. This model holds particular promise for patients with hemophilia-B who possess inhibitory antibodies against FIX. Our findings highlight the substantial potential of the MICE strategy as a gene replacement approach for treating genetic

disorders. Moreover, this approach opens up new possibilities for precise and targeted gene expression in other gene pairs involved in shared biological processes or pathways.

The present study aimed to optimize the HDR (homology-directed repair)-mediated gene editing strategy at the ITGA2B locus. To accomplish this, a Cas9-gRNA RNP complex was employed to introduce a double-strand break, followed by the transduction of cells with AAV6 carrying the donor template. Several donor template constructs, including AAV6-EF1α-GFP, AAV6-SA-P2A-GFP, and AAV6-IRES-GFP, were meticulously optimized to achieve monoallelic integration while preventing biallelic integration. This careful approach ensured coordinated and synchronized expression of both the integrated gene and the host gene. The study observed a positive correlation between HDR efficiency and the exponential increase in titre for the AAV6-EF1α-GFP donor. Notably, knock-in efficiencies surpassing 50% resulted in a decrease in ITGA2B (CD41a) expression, indicating biallelic integration of the transgene. Compared to other donors, the efficiency of SA-P2A-GFP and IRES-GFP integration was lower, with SA-P2A showing even lower efficiency than IRES-GFP. The reduced efficiency observed in SA-P2A-GFP may be potentially associated with the influence of the P2A sequence, as suggested by previous research (Szymczak-Workman *et al*., 2012). However, the specific reason for the limited integration, despite comparable viral titers to other expression components like IRES and EF1α promoter, remains unknown. Therefore, meticulous optimization of the P2A sequence in AAV vectors is essential to enhance efficiency. In the case of IRES-GFP, an interesting phenomenon was observed, where the integration of the IRES-GFP donor within intron 1 of the ITGA2B gene resulted in strong GFP expression without being removed during splicing. The precise mechanism by which translation initiation occurs from an intronic IRES is still not completely understood and likely involves intricate interactions among the IRES

element, ribosomes, and splicing machinery. It is speculated that the IRES could initiate translation at the pre-mRNA stage or even simultaneously with the splicing process. This discovery opens up new possibilities for achieving protein expression from pre-mRNA molecules, offering exciting avenues for future research.

In order to optimize HDR-mediated gene knock-in efficiency under the ITGA2B promoter and address potential toxicity concerns, the study initially focused on optimizing a promoterless donor template for expression under the ITGA2B promoter in HEL cell lines. Once the strategy was optimized, it was applied to CD34+ cells with the capacity to differentiate into megakaryocytes. HDR was carried out using SA-P2A-GFP and IRES-GFP donors, resulting in optimal expression of CD41a during megakaryocytic differentiation. However, HDR edited sample and virus alone treated control exhibited lower levels of megakaryopoiesis and viability, indicating toxicity associated with viral transduction. To mitigate these issues, a temporal validation of the transduction procedure was performed using the AAV6-IRES-GFP donor. The results showed that the efficiency of HDR reached a plateau after 12 hours of viral transduction. Removing the virus after 12 hours partially improved viability and cell count, but significantly enhanced megakaryopoiesis while maintaining HDR efficiency. It was observed that CD34+ cells had lower HDR efficiency compared to HEL cells, which could potentially be attributed to inherent characteristics specific to the cell type. These findings highlight the need to carefully consider and optimize the transduction procedure to minimize toxicity and maximize HDR efficiency, particularly when working with CD34+ cells. To address the issue of cell heterogeneity within the CD34+ population, one approach is to synchronize the cell cycle using molecular agents, which has the potential to enhance editing efficiency (Lin *et al*., 2014).

Further, the investigation focused on assessing the effects of gene editing on megakaryopoiesis and platelet biogenesis in HSPCs. To mitigate associated toxicity, CD34+ cells were employed at an increased cell density to alleviate stress induced by editing techniques like electroporation, transduction, and HDR pathway-related stress (Rogers *et al*., 2021). The relationship between the HR pathway and the apoptotic pathway has been well-established in previous studies, with Caspase-2 playing a crucial role as a mediator. Caspase-2 contributes to the maintenance of genomic stability, regulation of the cell cycle, and facilitation of DNA repair within the HR pathway. Acting as a tumor suppressor, Caspase-2 eliminates cells with DNA damage, promotes repair through the HR mechanism, and regulates cell cycle checkpoints. The diverse functions of Caspase-2 highlight its significance in preserving the integrity of the genome (Vigneswara and Ahmed, 2020). Flow cytometry analysis revealed notable enhancements in megakaryopoiesis within the edited samples, as indicated by elevated expression levels of CD41a and CD41a+CD42b+ markers in comparison to the unedited control. Ploidy analysis confirmed normal maturation of megakaryocytes, while the presence of proplatelets indicated imminent platelet production. Functional analysis further validated the successful generation of functional platelets derived from *In Vitro* megakaryopoiesis. However, despite the increased cell density, a decrease in both cell viability and cell count was observed, likely attributable to the gene editing process. Addressing the challenges associated with viability and cell count in gene editing studies requires further investigation to evaluate potential strategies for inhibiting cell death and improving delivery methods. Two promising strategies that could be explored include the utilization of a pan caspase inhibitor and a peptide-based Cas9 delivery system. The pan caspase inhibitor acts by inhibiting multiple caspase isoforms involved in programmed cell death pathways, thereby preventing cell death (Wang *et al*., 2020). On

the other hand, the peptide-based Cas9 delivery system enhances intracellular delivery of the Cas9 protein, facilitating efficient and stress-free gene editing (Foss *et al*., 2023). The combination of these strategies holds potential for mitigating cell death and improving overall cell count and viability during gene editing experiments. Nevertheless, comprehensive research is necessary to fully assess the effectiveness of these approaches.

One limitation encountered in this study is the difficulty in distinguishing mature megakaryocytes from platelets due to their shared expression of CD41a and CD42a/b markers. While size and granularity can serve as distinguishing factors, with platelets being smaller and exhibiting reduced granularity compared to megakaryocytes, it is important to note that platelets generated *In Vitro* during megakaryopoiesis may not precisely resemble those produced in vivo in terms of shape and size. To address this limitation and enhance the characterization of platelets, future studies should consider employing ex vivo models instead of relying solely on *In Vitro* models. One potential strategy involves the utilization of an immunodeficient FIX knockout mouse model. By introducing edited human CD34 cells into NGSGW mice with a FIX knockout, it becomes possible to generate human platelets from the edited HSPCs within the mouse model. After 16 weeks, administration of intraperitoneal clodronate liposomes can deplete mouse macrophages, thereby enriching human platelets in the circulation (Joo *et al*., 2022). This approach enables subsequent analysis and investigation of the characteristics and functionalities of the engineered platelets. To assess the functionality of the engineered platelets, assays such as tail clipping can be employed to measure their clotting ability (Zhang *et al*., 2010). The human platelets expressed in the mouse system can be separated as platelet-rich plasma (PRP), and assays such as platelet activation can accurately demonstrate the quantity of FIX carried by the platelets (Shi *et al*., 2014). This approach offers a more accurate representation of platelet behavior and

functionality compared to in vitro-derived platelets. The utilization of humanized FIX mice bridges the gap between *In Vitro* studies and clinical trials, facilitating the preclinical evaluation of platelet-targeted therapies. It enables assessments of safety, efficacy, and potential side effects associated with these therapies.

In summary, this study introduces MICE (Monoallelic Integration and Concerted Expression) as an innovative gene replacement strategy to overcome limitations in traditional HDR-based treatments for genetic disorders. MICE aims to achieve synchronized gene expression by precisely inserting a corrected gene sequence into a related gene locus, while preserving the expression of the host gene. By utilizing the ITGA2B promoter, the study successfully integrated the corrected FIX coding sequence, resulting in coordinated expression of coagulation GFP/FIX and ITGA2B specifically in the megakaryocytic lineage, leading to the generation of engineered platelets. This gene replacement strategy holds substantial potential for addressing the needs of hemophilia patients, particularly those with inhibitory antibodies against FIX, as it enables the packaging of FIX within the engineered platelets. The study conducted optimizations of the homology-directed repair (HDR)-mediated gene editing technique, emphasizing the importance of donor template constructs in achieving efficient monoallelic integration. Notably, these optimization efforts resulted in over 20% expression of GFP and FIX in platelets derived from edited HSPCs, while preserving ITGA2B expression. Furthermore, the study addressed the issue of defective megakaryopoiesis associated with viral transduction by implementing temporal changes in the transduction procedure. However, challenges related to toxicity in gene editing still remain, requiring further mitigation to improve cell viability and count. The study also discussed the challenge of distinguishing mature megakaryocytes from platelets and proposed the use of humanized FIX knockout mice as a valuable model for functionally validating platelets derived from transplanted human HSPCs. These mice

can serve as an effective platform for assessing the characteristics and functionality of platelets derived from edited HSPCs, bridging the gap between *In Vitro* studies and clinical trials. Overall, these findings underscore the significant potential of MICE as a gene replacement strategy and provide opportunities for achieving precise and targeted gene expression in shared biological processes or pathways.

SUMMARY AND CONCLUSION

This thesis aimed to optimize CRISPR/Cas9-based genome editing techniques, specifically base editing (BE) and homology-directed repair (HDR), in human hematopoietic stem and progenitor cells (HSPCs) for therapeutic applications in monogenic disorders. The overarching goal was to enhance the precision and efficacy of gene editing, validate functional outcomes, investigate underlying mechanisms, and address safety considerations.

The first study focused on harnessing CRISPR base editing to stimulate the expression of fetal hemoglobin (HbF) in erythroid progenitor cells and HSPCs as a potential therapeutic intervention. By utilizing adenine and cytosine base editors, we precisely targeted the highly homologous HBG promoter region, crucial for hemoglobin switching, while minimizing undesired genetic alterations. Notably, our investigations identified novel regulatory regions, including a cluster at the –123 position, and demonstrated that base editing at these sites led to higher levels of HbF expression compared to disruption of the well-established BCL11A binding site. Remarkably, HPFH-like mutations at the –123 and –124 positions resulted in the creation of a de novo binding site for KLF1, thereby promoting gamma-globin expression. These findings expand our understanding of the regulatory elements governing the HBG promoter and provide additional avenues for therapeutic upregulation of HbF. Collectively, our results underscore the potential of CRISPR base editing as a viable strategy for addressing monogenic disorders, particularly beta-hemoglobinopathies, and pave the way for further advancements in this field.

The second study introduced Monoallelic Integration and Concerted Expression (MICE) as an innovative gene replacement approach, specifically designed to overcome

limitations associated with traditional HDR-based therapies for genetic disorders. MICE enables synchronized gene expression by precisely inserting a corrected gene sequence into a related gene locus while preserving the expression of the host gene. Through meticulous optimization of HDR-mediated gene editing and leveraging the regulatory capabilities of the ITGA2B promoter, our investigation achieved coordinated expression of GFP/FIX and ITGA2B specifically within the megakaryocytic lineage, leading to the generation of engineered platelets. This novel strategy holds considerable promise for individuals with hemophilia, as it facilitates the packaging of GFP/FIX within platelets, enhancing therapeutic outcomes. Our optimization efforts yielded significant expression of both GFP and FIX in edited platelets while ensuring the maintenance of ITGA2B expression. Furthermore, we successfully addressed challenges related to toxicity and the distinction between mature megakaryocytes and platelets. To validate functionality, we proposed the utilization of humanized FIX knockout mice as an invaluable model. These significant findings highlight the potential of MICE as a precise gene replacement strategy, offering targeted gene expression within shared biological processes or pathways. This research opens up new avenues for further exploration and future clinical applications in the realm of genetic disorders.

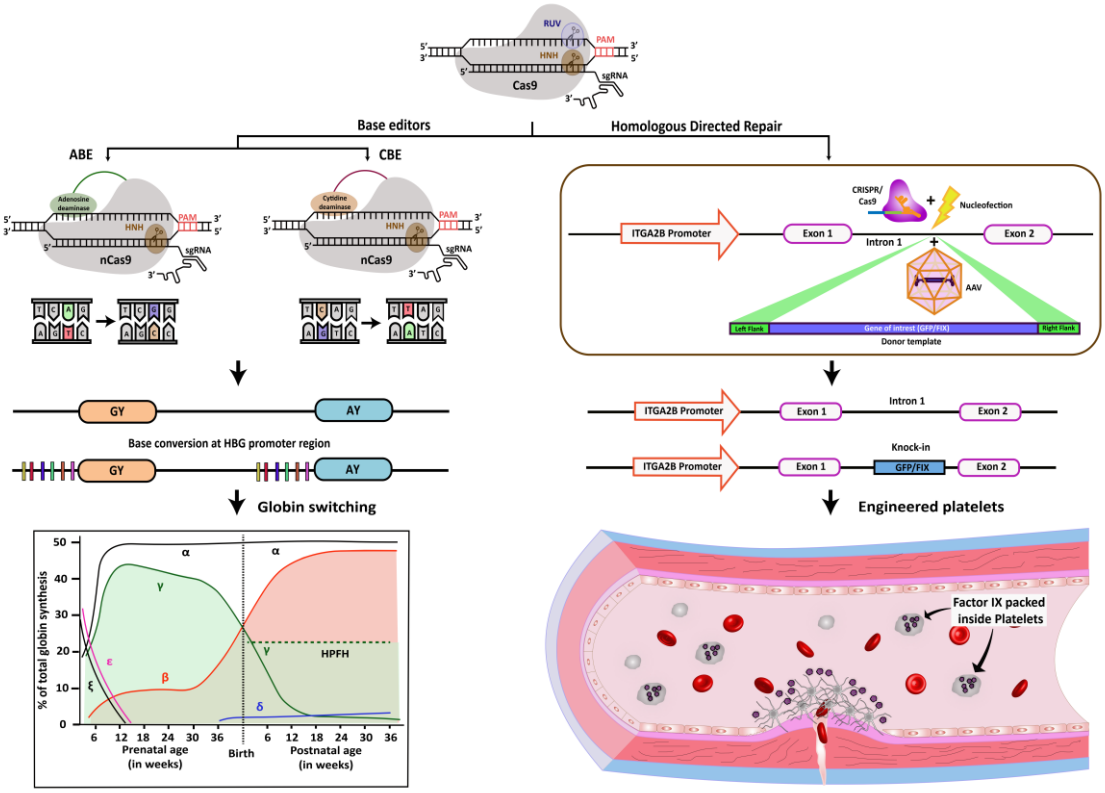


Figure 6.1.1: Schematic representation of a CRISPR/Cas9-based genome editing strategy involving base editing and HDR. Base editors were utilized to introduce HPFH-like point mutations in the HBG promoter, aiming to stimulate the production of HbF as a complement to the defective HbA globin in patients with β-hemoglobinopathies. In the case of HDR, the mutated FIX CDS was selectively replaced with the correct sequence within the ITGA2B promoter in a monoallelic manner, enabling coordinated expression of both genes to generate engineered platelets enriched with FIX for therapeutic applications in patients with hemophilia B.

FUTURE DIRECTIONS

Study 1:

1) Although the mechanism behind the newly identified target, involving the creation of a KLF1 binding site that regulates HBG expression, has been elucidated through EMSA, further investigations are required to understand the comprehensive regulatory pathway involved.
2) Base editing techniques should be applied to patients with sickle cell disease (SCD) or beta-thalassemia to assess their effectiveness and validate their potential as therapeutic interventions.
3) Preclinical studies, adhering to Good Manufacturing Practice (GMP) standards, must be conducted to confirm the safety and efficacy of the target before advancing to clinical trials.

Study 2:

1) The Monoallelic Integration and Concerted Expression (MICE) strategy, utilizing homology-directed repair (HDR), implemented in this study, should be expanded to encompass additional genes that are involved in the same biological pathway. This extension will facilitate a more comprehensive exploration of its applicability and elucidate its potential impact on gene expression dynamics within the pathway.
2) Investigations should be carried out using humanized immunocompromised mouse models to obtain platelets that closely resemble human platelets, enabling a more accurate validation of the therapeutic effects of FIX in platelets.

3) Mice studies should also be conducted to evaluate the engraftment potential and multilineage differentiation capacity of the edited cells, ensuring the safety and stability of the edited genome.
4) Additionally, studies should explore the use of small molecules that inhibit apoptosis and develop new delivery methods to enhance the viability and cell count of edited cells.

BIBLIOGRAPHY

Adli M (2018) The CRISPR tool kit for genome editing and beyond. *Nature Communications* 9(1). Springer US. DOI: 10.1038/s41467-018-04252-2.

Al-Saif AM (2019) Gene therapy of hematological disorders: current challenges. *Gene Therapy* 26(7–8). Springer US: 296–307. DOI: 10.1038/s41434-019-0093-4.

Amend SR, Valkenburg KC and Pienta KJ (2016) Murine hind limb long bone dissection and bone marrow isolation. *Journal of Visualized Experiments* 2016(110): 3–6. DOI: 10.3791/53936.

Andrew V. Anzalone LWK and DRL (2020) Genome editing with CRISPR-Cas nucleases, base editors , transposases and prime editors. Nat. Biotechnology.

Ann Ran *et al* 2013 (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6): 1380–1389. DOI: 10.1016/j.cell.2013.08.021.Double.

Anzalone A V., Randolph PB, Davis JR, *et al*. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576(7785). Springer US: 149–157. DOI: 10.1038/s41586-019-1711-4.

Arbab M, Shen MW, Mok B, *et al*. (2020) Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* 182(2): 463-480.e30. DOI: 10.1016/j.cell.2020.05.037.

Azhagiri MKK, Babu P, Venkatesan V, *et al*. (2021) Homology-directed gene-editing approaches for hematopoietic stem and progenitor cell gene therapy. *Stem Cell Research and Therapy* 12(1). BioMed Central: 1–12. DOI: 10.1186/s13287-021-02565-6.

Bae S, Park J and Kim JS (2014) Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30(10): 1473–1475. DOI: 10.1093/bioinformatics/btu048.

Bae S, Kweon J, Kim HS, *et al*. (2014) Microhomology-based choice of Cas9 nuclease target sites. *Nature Methods* 11(7). Nature Publishing Group: 705–706. DOI: 10.1038/nmeth.3015.

Bak RO, Dever DP, Reinisch A, *et al*. (2017) Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *eLife* 6: 1–19. DOI: 10.7554/eLife.27873.

Bak RO, Dever DP and Porteus MH (2018) CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nature Protocols* 13(2): 358–376. DOI: 10.1038/nprot.2017.143.

Barbarani G, Łabedz A and Ronchi AE (2020) β-Hemoglobinopathies: The Test Bench for Genome Editing-Based Therapeutic Strategies. *Frontiers in Genome Editing* 2(December): 1–10. DOI: 10.3389/fgeed.2020.571239.

Barczak W, Suchorska W, Rubiś B, *et al*. (2014) Universal Real-Time PCR-Based Assay for Lentiviral Titration. *Molecular Biotechnology* 57(2): 195–200. DOI: 10.1007/s12033-014-9815-4.

Batista Napotnik T, Polajžer T and Miklavčič D (2021) Cell death due to electroporation – A review. *Bioelectrochemistry* 141. DOI: 10.1016/j.bioelechem.2021.107871.

Batty P and Lillicrap D (2019) Advances and challenges for hemophilia gene therapy. *Human Molecular Genetics* 28(R1): R95–R101. DOI: 10.1093/hmg/ddz157.

Batty P and Lillicrap D (2021) Gene therapy for hemophilia: Current status and laboratory consequences. *International Journal of Laboratory Hematology* 43(S1): 117–123. DOI: 10.1111/ijlh.13605.

Berntorp E, Fischer K, Hart DP, *et al*. (2021) Haemophilia. *Nature Reviews Disease Primers* 7(1). DOI: 10.1038/s41572-021-00278-x.

Bijlani S, Pang KM, Sivanandam V, *et al*. (2022) The Role of Recombinant AAV in Precise Genome Editing. *Frontiers in Genome Editing* 3(January): 1–16. DOI: 10.3389/fgeed.2021.799722.

Bloomer H, Smith RH, Hakami W, *et al*. (2021) Genome editing in human hematopoietic stem and progenitor cells via CRISPR-Cas9-mediated homology-independent targeted integration. *Molecular Therapy* 29(4). Elsevier Ltd.: 1611–1624. DOI: 10.1016/j.ymthe.2020.12.010.

Clement K, Rees H, Canver Matthew C, *et al*. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* 2. Springer US: 10–12. DOI: 10.1038/s41587-019-0032-3.

Corn JE (2017) Preparation of PCR amplicons from edited cells for deep sequencing. *https://www.protocols.io/view/preparation-of-pcr-amplicons-from-edited-cells-for-6ruhd6w*. Available at: https://www.protocols.io/view/preparation-of-pcr-amplicons-from-edited-cells-for-6ruhd6w.

Crossley M, Whitelaw E, Perkins A, *et al*. (1996) Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. *Molecular and Cellular Biology* 16(4): 1695–1705. DOI: 10.1128/mcb.16.4.1695.

Cullot G, Boutin J, Toutain J, *et al*. (2019) CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature Communications* 10(1). Springer US: 1–14. DOI: 10.1038/s41467-019-09006-2.

Daniel E. Bauer and Stuart H.Orkin (2012) Update on fetal hemoglobin gene regulation in hemoglobinopathies. *Curr Opin Pediatr* 23(1): 617–632. DOI: 10.1097/MOP.0b013e3283420fd0.Update.

Ewels PA, Peltzer A, Fillinger S, *et al*. (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* 38(3): 276–278. DOI: 10.1038/s41587-020-0439-x.

Foss D V., Muldoon JJ, Nguyen DN, *et al*. (2023) Peptide-mediated delivery of CRISPR enzymes for the efficient editing of primary human lymphocytes. *Nature Biomedical Engineering* 7(May). Springer US. DOI: 10.1038/s41551-023-01032-2.

Franchini M and Mannucci PM (2011) Inhibitors of propagation of coagulation (factors VIII, IX and XI): A review of current therapeutic practice. *British Journal of Clinical Pharmacology* 72(4): 553–562. DOI: 10.1111/j.1365-2125.2010.03899.x.

Frangoul H, Altshuler D, Cappellini MD, *et al*. (2021) CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine* 384(3): 252–260. DOI: 10.1056/nejmoa2031054.

Frati G and Miccio A (2021) Genome editing for β-hemoglobinopathies: Advances and challenges. *Journal of Clinical Medicine* 10(3): 1–17. DOI: 10.3390/jcm10030482.

Fucharoen S, Shimizu K and Fukumaki Y (1990) A novel C-T transition within the distal CCAAT motif of the G ’ y-globin gene in the Japanese HPFH : implication of factor binding in elevated fetal globin expression. *Nucleic Acids Research* 18(17): 5245–5253.

Gaj T, Staahl BT, Rodrigues GMC, *et al*. (2017) Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery. *Nucleic Acids Research* 45(11): 1–11. DOI: 10.1093/nar/gkx154.

Gaudelli Nicole M, Komor AC, Rees HA, *et al*. (2017) Programmable base editing of A • T to G • C in genomic DNA without DNA cleavage. *Nature Publishing Group* 551(7681). Nature Publishing Group: 464–471. DOI: 10.1038/nature24644.

Gaudelli Nicole M., Komor AC, Rees HA, *et al*. (2017) Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* 551(7681). Nature Publishing Group: 464–471. DOI: 10.1038/nature24644.

Genovese P, Schiroli G, Escobar G, *et al*. (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* 510(7504): 235–240. DOI: 10.1038/nature13420.

Heckl D, Kowalczyk MS, Yudovich D, *et al*. (2014) Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnology* 32(9): 941–946. DOI: 10.1038/nbt.2951.

Hsiau T, Maures T, Waite K, *et al*. (2018) Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*. Available at: https://doi.org/10.1101/251082.

Ikuta T, Papayannopoulou T, Stamatoyannopoulos G, *et al*. (1996) Globin Gene Switching. *The Journal of Biological Chemistry* 271(24): 14082–14091.

J.G.Gilman *et al* 1988 (1988) Distal CCAAT box deletion in the Aγ globin gene of two black adolescents with elevated fetal Aγ globin. *Nucleic Acids Research* 16(22): 10635–10642.

James T. Robinson *et al* 2012 (2012) Integrative Genomics Viewer. *Nat Biotechnol* 29(1): 24–26. DOI: 10.1038/nbt.1754.Integrative.

Joo JH, Wang X, Singh S, *et al*. (2022) Intraosseous delivery of platelet-targeted factor VIII lentiviral vector in humanized NBSGW mice. *Blood Advances* 6(19). The American Society of Hematology: 5556–5569. DOI: 10.1182/bloodadvances.2022008079.

Kim YB, Komor AC, Levy JM, *et al*. (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology* 35(4). Nature Publishing Group: 371–376. DOI: 10.1038/nbt.3803.

Kluesner MG, Nedveck DA, Lahr WS, *et al*. (2018) EditR : A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal* 1(3): 239–250. DOI: 10.1089/crispr.2018.0014.

Koblan LW, Doman JL, Wilson C, *et al*. (2018) Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Publishing Group* (May). Nature Publishing Group. DOI: 10.1038/nbt.4172.

Koblan LW, Arbab M, Shen MW, *et al*. (2021) Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nature Biotechnology* 39(11). Springer US: 1414–1425. DOI: 10.1038/s41587-021-00938-z.

Koblan LW, Erdos MR, Wilson C, *et al*. (2021) In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature* 589(7843). Springer US: 608–614. DOI: 10.1038/s41586-020-03086-7.

Komor AC, Kim YB, Packer MS, Zuris JA and David R (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533(7603): 420–424. DOI: 10.1038/nature17946.Programmable.

Kosicki M, Allen F, Steward F, *et al*. (2022) Cas9-induced large deletions and small indels are controlled in a convergent fashion. *Nature Communications* 13(1). Springer US: 1–11. DOI: 10.1038/s41467-022-30480-8.

Krejci L, Altmannova V, Spirek M, *et al*. (2012) Homologous recombination and its regulation. *Nucleic Acids Research* 40(13): 5795–5818. DOI: 10.1093/nar/gks270.

Kurita R, Suda N, Sudo K, *et al*. (2013) Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLoS ONE* 8(3): 1–15. DOI: 10.1371/journal.pone.0059890.

Kuzminov A (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America* 98(15): 8241–8246. DOI: 10.1073/pnas.131009198.

Kyung-Rok Yu *et al* 2016 (2016) Targeted Genome Editing III The Cytotoxic Effect of RNA-Guided Endonuclease Cas9 on Human Hematopoietic Stem and Progenitor Cells (HSPCs). *Molecular Therapy* 24(May). The American Society of Gene & Cell Therapy: S225–S226. DOI: 10.1016/S1525-0016(16)33372-X.

Labun K, Montague TG, Krause M, *et al*. (2019) CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Research* 47(W1): W171–W174. DOI: 10.1093/nar/gkz365.

Lam DK, Feliciano PR, Arif A, *et al*. (2023) Improved cytosine base editors generated from TadA variants. *Nature Biotechnology*. Springer US. DOI: 10.1038/s41587-022-01611-9.

Langmead B and Salzberg S (2013) Fast gapped-read alignment with Bowtie 2 Ben. *Nature methods* 9(4): 357–359. DOI: 10.1038/nmeth.1923.Fast.

Lee ABC, Tan M-H and Chai CLL (2022) Small-molecule enhancers of CRISPR-induced homology-directed repair in gene therapy: A medicinal chemist's perspective. *Drug Discovery Today* 27(9): 2510–2525. DOI: https://doi.org/10.1016/j.drudis.2022.06.006.

Lee HK, Willi M, Miller SM, *et al*. (2018) Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nature Communications* 9(1). Springer US: 7–12. DOI: 10.1038/s41467-018-07322-7.

Leebeek FWG and Miesbach W (2021) Gene therapy for hemophilia: a review on clinical benefit, limitations, and remaining issues. *Blood* 138(11): 923–931. DOI: 10.1182/blood.2019003777.

Leonard A, Tisdale JF and Bonner M (2022) Gene Therapy for Hemoglobinopathies: Beta-Thalassemia, Sickle Cell Disease. *Hematology/Oncology Clinics of North America* 36(4): 769–795. DOI: 10.1016/j.hoc.2022.03.008.

Li C, Psatha N, Sova P, *et al*. (2018) Reactivation of g-globin in adult b-YAC mice after ex vivo and in vivo hematopoietic stem cell genome editing. *Blood* 131(26): 2915–2928. DOI: 10.1182/blood-2018-03-838540.

Li X and Heyer WD (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell Research* 18(1): 99–113. DOI: 10.1038/cr.2008.1.

Li ZH, Wang Jun, Xu JP, *et al*. (2023) Recent advances in CRISPR-based genome editing technology and its applications in cardiovascular research. *Military Medical Research* 10(1). BioMed Central: 1–20. DOI: 10.1186/s40779-023-00447-x.

Lin S, Staahl BT, Alla RK, *et al*. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3: e04766. DOI: 10.7554/eLife.04766.

Liu HAR and DR (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet.* 19(12): 770–788. DOI: 10.1038/s41576-018-0059-1.Base.

Liu N, Hargreaves V V, Zhu Q, *et al*. (2018) Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch Article Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell* 173. Elsevier Inc.: 1–13. DOI: 10.1016/j.cell.2018.03.016.

Locatelli F, Thompson AA, Kwiatkowski JL, *et al*. (2022) Betibeglogene Autotemcel Gene Therapy for Non–β 0 /β 0 Genotype β-Thalassemia . *New England Journal of Medicine* 386(5): 415–427. DOI: 10.1056/nejmoa2113206.

Loucari CC, Patsali P, Van Dijk TB, *et al*. (2018) Rapid and Sensitive Assessment of Globin Chains for Gene and Cell Therapy of Hemoglobinopathies. *Human Gene Therapy Methods* 29(1): 60–74. DOI: 10.1089/hgtb.2017.190.

M´etais J-Y, Doerfler PA, Mayuranathan T, *et al*. (2019) Genome editing of HBG1 and HBG2 to induce fetal hemoglobin. *Blood Advance* 3(21): 3379–3392. DOI: 10.1182/bloodadvances.2019000820.

Mahalingam G, Mohan A, Arjunan P, *et al*. (2022) Using Lipid Nanoparticles for the Delivery of Chemically Modified mRNA into Mammalian Cells. *JoVE*. MyJoVE Corp: e62407. DOI: doi:10.3791/62407.

Mark A. DeWitt WM et al. (2016) Selection-free Genome Editing of the Sickle Mutation in Human Adult Hematopoietic Stem/Progenitor Cells. *Sci Transl Med* 8(360). DOI: 10.1126/scitranslmed.aaf9336.Selection-free.

Martyn GE, Wienert B, Yang L, *et al*. (2018) Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nature Genetics* 50. Springer US: 498–503. DOI: 10.1038/s41588-018-0085-0.

Martyn GE, Wienert B, Kurita R, *et al*. (2019) A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. *Blood* 133(8): 852–856. DOI: 10.1182/blood-2018-07-863951.

Matsunaga T, Tanaka I, Kobune M, *et al*. (2006) Ex Vivo Large-Scale Generation of Human Platelets from Cord Blood CD34 + Cells . *Stem Cells* 24(12): 2877–2887. DOI: 10.1634/stemcells.2006-0309.

McIntosh BE, Brown ME, Duffin BM, *et al*. (2015) Nonirradiated NOD,B6.SCID Il2rγ-/- kitW41/W41 (NBSGW) mice support multilineage engraftment of human hematopoietic cells. *Stem Cell Reports* 4(2). The Authors: 171–180. DOI: 10.1016/j.stemcr.2014.12.005.

Miller IJ and Bieker JJ (1993) A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Molecular and Cellular Biology* 13(5): 2776–2786. DOI: 10.1128/mcb.13.5.2776.

Monian P, Shivalila C, Lu G, *et al*. (2022) Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides. *Nature Biotechnology* 40(7): 1093–1102. DOI: 10.1038/s41587-022-01225-1.

Motum *et al* 1994 (1994) The Australian type of nondeletional Gy-HPFH has a C+G substitution at nucleotide - 114 of the Gy gene. *British journal of Haematology* 86: 219–221.

Neugebauer ME, Hsu A, Arbab M, *et al*. (2022) Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nature Biotechnology*. Springer US. DOI: 10.1038/s41587-022-01533-6.

Perdomo J, Yan F, Leung HHL, *et al*. (2017) Megakaryocyte differentiation and platelet formation from human cord blood-derived CD34+ cells. *Journal of Visualized Experiments* 2017(130): 1–8. DOI: 10.3791/56420.

Piras BA, Drury JE, Morton CL, *et al*. (2016) Distribution of AAV8 particles in cell lysates and culture media changes with time and is dependent on the recombinant vector. *Molecular Therapy - Methods and Clinical Development* 3(November 2015). Official journal of the American Society of Gene & Cell Therapy: 16015. DOI: 10.1038/mtm.2016.15.

Porteus MH (2019) A New Class of Medicines through DNA Editing. *New England Journal of Medicine* 380(10): 947–959. DOI: 10.1056/nejmra1800729.

Prasad K, George A, Ravi NS, *et al*. (2021) CRISPR/Cas based gene editing: marking a new era in medical science. *Molecular Biology Reports* 48(5). Springer Netherlands: 4879–4895. DOI: 10.1007/s11033-021-06479-7.

Psatha N, Reik A, Phelps S, *et al*. (2018) Disruption of the BCL11A Erythroid Enhancer Reactivates Fetal Hemoglobin in Erythroid Cells of Patients with β-Thalassemia Major. *Molecular Therapy - Methods and Clinical Development* 10(September). Elsevier Ltd.: 313–326. DOI: 10.1016/j.omtm.2018.08.003.

Ran FA, Hsu PD, Wright J, *et al*. (2013) Genome engineering using the CRISPR-Cas9 system. *NATURE PROTOCOLS* 8(11): 2281–2308. DOI: 10.1038/nprot.2013.143.

Rashighi M and Harris JE (2017) 乳鼠心肌提取 HHS Public Access. *Physiology & behavior* 176(3): 139–148. DOI: 10.1053/j.gastro.2016.08.014.CagY.

Ravi NS, Wienert B, Wyman SK, *et al*. (2022) Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin. *eLife* 11. DOI: 10.7554/ELIFE.65421.

Rees DC, Williams TN and Gladwin MT (2010) Sickle-cell disease. *The Lancet* 376(9757). Elsevier Ltd: 2018–2031. DOI: 10.1016/S0140-6736(10)61029-X.

Richardson CD, Ray GJ, DeWitt MA, *et al*. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.

Nature Biotechnology 34(3). Nature Publishing Group: 339–344. DOI: 10.1038/nbt.3481.

Richter MF, Zhao KT, Eton E, *et al*. (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology* 38(7). Springer US: 883–891. DOI: 10.1038/s41587-020-0453-z.

Rogers GL, Huang C, Clark RDE, *et al*. (2021) Optimization of AAV6 transduction enhances site-specific genome editing of primary human lymphocytes. *Molecular Therapy - Methods and Clinical Development* 23(December). Elsevier Ltd.: 198–209. DOI: 10.1016/j.omtm.2021.09.003.

S. Zertal-Zidani, Merghoub T, Ducrocq R, Inserm U, *et al*. (1999) A Novel C>A Transversion Within the Distal CCAAT Motif of the Gy-Globin Gene in the Algerian Gy-beta+ - Hereditary Persistence of Fetal Hemoglobin. *Hemoglobin* 23(2): 159–169.

Sambrook J and Russell DW (2001) Preparation and Transformation of Competent E . coli Using Calcium Chloride. *Cold Spring Harbor Protocols*: 1–2.

Sanjana NE, Shalem O and Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening HHS Public Access Supplementary Material. *Nat Methods* 11(8): 783–784. DOI: 10.1038/nmeth.3047.Improved.

Schiml S, Fauser F and Puchta H (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. *Proceedings of the National Academy of Sciences of the United States of America* 113(26): 7266–7271. DOI: 10.1073/pnas.1603823113.

Shakirova A, Karpov T, Komarova Y, *et al*. (2023) In search of an ideal template for therapeutic genome editing: A review of current developments for structure optimization. *Frontiers in Genome Editing* 5(February): 1–19. DOI: 10.3389/fgeed.2023.1068637.

Shalem O, Sanjana NE, Hartenian E, *et al*. (2014) Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* 343(6166): 84–87. DOI: 10.1126/science.1247005.Genome-Scale.

Sharma R, Anguela XM, Doyon Y, *et al*. (2016) In vivo genome editing of the albumin locus as a platform for protein replacement therapy. 126(15): 1777–1785. DOI: 10.1182/blood-2014-12-615492.

Shi Q, Kuether EL, Chen Y, *et al*. (2014) Platelet gene therapy corrects the hemophilic phenotype in immunocompromised hemophilia A mice transplanted with genetically manipulated human cord blood stem cells. *Blood* 123(3): 395–403. DOI: 10.1182/blood-2013-08-520478.

Stoming TA *et al* (1989) An A gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T–C mutation at position -175 to the cap site of the A gamma globin gene. *Blood* 329: 329–333.

Stone A V, Vanderman KS, Willey JS, *et al*. (2014) Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnology* 32(10): 941–946. DOI: 10.1038/nbt.2951.Generation.

Sung P and Klein H (2006) Mechanism of homologous recombination: Mediators and helicases take on regulatory functions. *Nature Reviews Molecular Cell Biology* 7(10): 739–750. DOI: 10.1038/nrm2008.

Suzuki K, Tsunekawa Y, Hernandez-Benitez R, *et al*. (2016) Homology-independent targeted integration. *Nature* 540(7631): 144–149. DOI: 10.1038/nature20565.In.

Suzuki K, Yamamoto M, Hernandez-Benitez R, *et al*. (2019) Precise in vivo genome editing via single homology arm donor mediated intron-targeting gene integration for genetic disease correction. *Cell Research* 29(10). Springer US: 804–819. DOI: 10.1038/s41422-019-0213-0.

Szymczak-Workman AL, Vignali KM and Vignali DAA (2012) Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harbor Protocols* 7(2): 199–204. DOI: 10.1101/pdb.ip067876.

Tao J, Bauer DE and Chiarle R (2023) Assessing and advancing the safety of CRISPR-Cas tools: from DNA to RNA editing. *Nature Communications* 14(1). Springer US. DOI: 10.1038/s41467-023-35886-6.

Trakarnsanga K, Griffiths RE, Wilson MC, *et al*. (2017) An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells. *Nature Communications* 8(May 2016). Nature Publishing Group. DOI: 10.1038/ncomms14750.

Traxler EA, Yao Y, Wang Y, *et al*. (2016) A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *NATURE MEDICINE* 22(9): 987–990. DOI: 10.1038/nm.4170.

Tröder SE and Zevnik B (2022) History of genome editing: From meganucleases to CRISPR. *Laboratory Animals* 56(1): 60–68. DOI: 10.1177/0023677221994613.

VE Tate WW and DW (1986) The British form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to an S1 hypersensitive site 5’ to the A gamma globin gene. *Blood* 68: 1389–1393. DOI: 10.1182/blood.V68.6.1389.bloodjournal6861389.

Venkatesan V, Srinivasan S, Babu P, *et al*. (2021) Manipulation of Developmental Gamma-Globin Gene Expression: an Approach for Healing Hemoglobinopathies. *Molecular and Cellular Biology* 41(1): 1–18. DOI: 10.1128/mcb.00253-20.

Vigneswara V and Ahmed Z (2020) The Role of Caspase-2 in Regulating Cell Fate. *Cells* 9(5). DOI: 10.3390/cells9051259.

Wang C, Chang CC, Wang L, *et al*. (2020) Inhibition of caspases improves non-viral t cell

receptor editing. *Cancers* 12(9): 1–15. DOI: 10.3390/cancers12092603.

Wang J, Exline CM, Declercq JJ, *et al*. (2015) Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nature Biotechnology* 33(12). Nature Publishing Group: 1256–1263. DOI: 10.1038/nbt.3408.

Webber BR, Lonetree C lin, Kluesner MG, *et al*. (2019) Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nature Communications* 10(1). Springer US. DOI: 10.1038/s41467-019-13007-6.

Wienert B, Funnell APW, Norton LJ, *et al*. (2015) Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nature Communications* 6(1). Nature Publishing Group: 1–8. DOI: 10.1038/ncomms8085.

Wienert B, Martyn GE, Kurita R, *et al*. (2017) KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood* 130(6): 803–807. DOI: 10.1182/blood-2017-02-767400.

Wienert B, Martyn GE, Funnell APW, *et al*. (2018) Wake-up Sleepy Gene : Reactivating Fetal Globin for b -Hemoglobinopathies. *Trends in Genetics* 34(12). Elsevier Ltd: 927–940. DOI: 10.1016/j.tig.2018.09.004.

Wimberly H, Shee C, Thornton PC, *et al*. (2013) R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nature Communications* 4. Nature Publishing Group: 1–10. DOI: 10.1038/ncomms3115.

Yip BH (2020) Recent advances in CRISPR/Cas9 delivery strategies. *Biomolecules* 10(6). DOI: 10.3390/biom10060839.

Yu KR, Natanson H and Dunbar CE (2016) Gene Editing of Human Hematopoietic Stem and Progenitor Cells: Promise and Potential Hurdles. *Human Gene Therapy* 27(10): 729–740. DOI: 10.1089/hum.2016.107.

Zafra MP, Schatoff EM, Katti A, *et al*. (2018) Optimized base editors enable efficient editing in cells, organoids and mice. *Nature Biotechnology* 36(9): 888–893. DOI: 10.1038/nbt.4194.

Zarif MN, Soleimani M, Abolghasemi H, *et al*. (2011) The high yield expansion and megakaryocytic differentiation of human umbilical cord blood CD133 + cells. *Cell Journal* 13(3): 173–178.

Zhang G, Shi Q, Fahs SA, *et al*. (2010) Factor IX ectopically expressed in platelets can be stored in α-granules and corrects the phenotype of hemophilia B mice. *Blood* 116(8): 1235–1243. DOI: 10.1182/blood-2009-11-255612.

Zhang X, Zhu B, Chen L, *et al*. (2020) Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nature Biotechnology* 38(7). Springer US: 856–860. DOI: 10.1038/s41587-020-0527-y.

Annexure

List of Publications

1. Ravi, N. S. *et al.* Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin. *Elife* **11**, (2022).
2. George, A. *et al.* Efficient and error-free correction of sickle mutation in human erythroid cells using prime editor-2. *Front. Genome Ed.* **4**, 1–16 (2022).
3. Prasad, K., George, A., Ravi, N. S. & Mohankumar, K. M. CRISPR/Cas based gene editing: marking a new era in medical science. *Mol. Biol. Rep.* **48**, 4879–4895 (2021).
4. Devaraju, N., Rajendiran, V., Ravi, N. S. & Mohankumar, K. M. Genome Engineering of Hematopoietic Stem Cells Using CRISPR/Cas9 System. *Methods Mol. Biol.* **2429**, 307–331 (2022).
5. Christopher, A. C. *et al.* Preferential Expansion of Human CD34+CD133+CD90+Hematopoietic Stem Cells Enhances Gene-Modified Cell Frequency for Gene Therapy. *Hum. Gene Ther.* **33**, 188–201 (2022).
6. Ravi,NS. *Et al*. Arrayed gRNA screening by base editors in mammalian cell lines using lentiviral system (in progress).
7. Ravi,NS. *Et al*. Monoallelic Integration and Concerted Expression (MICE): A strategy for therapeutic gene replacement (in progress).

Curriculum Vitae

R. Nithin Sam
¾ Jones Street, Nagercoil,
Kanyakumari district, Tamil Nadu – 629001
Mobile: +91 9489281860
Email Id: nithinsam91@gmail.com, nithinsam91@cmcvellore.ac.in

Research experience:

2017 - Till date

PhD - Senior Research Fellow (SRF)
Institute: **Centre for Stem Cell Research (a unit of inStem),** Christian Medical College (CMC), Vellore, TamilNadu, India and **Sree Chitra Tirunal Institute** for Medical Sciences and Technology, Thiruvananthapuram, Kerala, India.
Project title: *Genome editing strategies for the treatment of hereditary hematological disorders; Hemophilia and β-thalassemia.*
Supervised by: **Dr. Mohankumar K. Murugesan**

Technical skills: Genome editing tools (Cas9 based knock out, Cas9 based HDR, Base editing), Gene therapy tools (Lentivirus and Adeno Associated Virus), T7 endonuclease assay, Invitro cas9 assay, mRNA synthesis for gene editing, Handling primary cells (healthy and patient HSPCs), Megakaryopoiesis and platelet production from CD34+ cells, Erythropoiesis from CD34+cells, Cell line handling (HEK 293T, HEL, HUDEP-2 and K562), Stable cell line production, Transfection, Nucleofection (Lonza, Maxcyte and Neon), ddPCR (sybr and probe), Protein purification, Molecular biology (gRNA design, basic cloning, oligo annealing, ligation, restriction digestion and HIFI assembly), qRT PCR (Vector copy number validation, AAV MOI evaluation, expression analysis and large deletion validation), mice studies (handled NSG and NBSGW strains, Chimerism studies) Basic bioinformatics (Ubuntu based tools), Western blot, SDS and Native page, Software (R studio, Prism, MS Office and Adobe illustrator).
Other Skills: Lab management, proposal writing and standardizations.

2015 – 2017

Research associate
Institute: **Department of Medical Microbiology**, Christian Medical College, Vellore, Tamil Nadu, India.
Project titles: *1) Expression study of efflux pump in extensively drug resistant Pseudomonas aeruginosa. 2) New method to find*

Minimum Biofilm Eradication Concentration (MBEC) for biofilm forming organism. 3) Outbreak surveillance of Acinetobacter baumannii using Multi Loci Sequence Typing. 4) A novel mutation in lipid-A biosynthesis pathway which leads to colistin resistance in Acinetobacter baumannii.
Supervised by: **Dr. Balaji Veraragavan**

Technical skills: qRT PCR, Next Generation Sequencing (ion torrent), Sanger sequencing, Epidemiological typing methods and Microbiology techniques.

Education:

2013 – 2015 **Post graduate program in M.Tech Medical Biotechnology**
Institute: **Dr. M.G.R. Educational and Research Institute**, Dr. M.G.R. University, Chennai, Tamil Nadu, India.
Thesis title: *Expression study of efflux pump in multidrug resistant Klebsiella pneumonia*
Supervised by: **Dr. Rama Vaidyanathan**

Technical skills: Microbiology techniques, RT-PCR, Real time PCR, Literature review and Bioinformatics tools.
Year of passing: 2015 OGPA: **9.28**

2009 – 2013 **Undergraduate program in B.Tech Biotechnology**
Institute: **Agricultural College and Research Institute**, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India.
Project title: *Harnessing host resistance against White Fly in Tomato*
Supervised by: **Dr. Murugan and Dr. Angapan**

Skills developed: Teamwork, Paper presentation, Poster presentation, Video presentation, Cloning, Plant cell culture and Bioinformatics tools.
Year of passing: 2013 OGPA: **8.24**

2008 – 2009 Higher secondary school education School: **C.S.I Matriculation Higher Secondary School**, Nagercoil, Kanyakumari district, Tamil Nadu, India.
Year of passing: 2009 Percentage: **82.58%**

2006 – 2007 Secondary school education School: **C.S.I Matriculation Higher Secondary School**, Nagercoil, Kanyakumari district, Tamil Nadu, India.
Year of passing: 2007 Percentage: **76.72%**

Experience:

2012 Internship in Molecular biology and Tissue culture,
Institution: **Institute of Agricultural Biotechnology**, University of agricultural sciences, Dharwad, Karnataka, India.
Project titles: *Transformation of tomato plants with endochitinase and β- glucanase gene from Trichoderma spp. by Agrobacterium mediated transformation and PCR screening for transformed plants.*
Time period: 4 months
Supervised by: **Dr. Sumangala Bhat**

Skill developed: Screening of Trichoderma spp. for antifungal activity against Sclerotium rolfsii and Rhizoctonia solani, cloning of sm1 gene from Trichoderma virens, Tri Parental Mating for Agrobacterim transformation, Primer designing, fungal handling techniques, monoclonal antibody production using phage display, cloning, transgenics and plant tissue culture.

Professional: Affiliations Reviewer in **Indian Journal of Medical Microbiology** from 2016-2017.

Awards: Poster presentation awards (Research Day 2018 and 2022, International Bone marrow symposium 2022)
Oral presentation awards (Research Day 2019)
Travel grant (International Travel Support (ITS) – SERB, 2019)
CSIR SRF direct, 2021
IIN Research Excellence Award, 2022

Conference Attended Cell Symposia (Gene and Cell Based Therapies: CRISPR, Stem Cells and Beyond, March 2-4, 2020 – San Francisco, CA, USA.
Cell and Gene Therapy Symposia (2017,2018, 2019, 2020, 2021, 2022), Center for Stem Cell Research, CMC, Vellore, India.
International Bone Marrow Symposium (2019, 2022)
Christian Medical College, Vellore, India.

Software Proficiency: Bioinformatics tools, Microsoft office, R Studio, Ubuntu based t tools and Adobe (Illustrator and Photoshop)

Additional Certifications: Completed Business English Certificate course (Cambridge university -ESOL, London).

Extra-Curricular Best volunteer in National Service Scheme, active participant in wild animal census, wildlife Photographer (Bird photography).

Activities:

Linguistic skills: Read (English, Tamil, Hindi), Write (English, Tamil, Hindi), Speak (English, Tamil, Malayalam).

Declaration:

I hereby declare that all the information furnished above is true to the best of my knowledge.

Appendices

Appendix a - Ethics committee approval

January 31, 2020.

Dr. Mohankumar K,
Assistant Investigator,
Department of CSCR,
Christian Medical College,
Vellore – 632 004

Sub: **External Research Grant: New Proposal:DBT**
Recreating HPFH-associated point mutation to elevate the fetal haemoglobin levels for the treatment of beta-hemoglobinopathies using CRISPR based tools.
Dr. Mohankumar K. Murugesan, Assistant Investigator, CSCR, Dr. Srujan Marepally, Dr. Alok Srivastava, CSCR & Haematology

Ref: IRB Min. No. 12309 (OTHER) dated 30.10.2019

Dear Dr. Mohankumar K,

The Institutional Review Board (**Silver**, Research and Ethics Committee) of the Christian Medical College, Vellore, reviewed and discussed your project titled "Recreating HPFH-associated point mutation to elevate the fetal haemoglobin levels for the treatment of beta-hemoglobinopathies using CRISPR based tools" on October 30, 2019. I am quoting below the minutes of the meeting.

The Committee reviewed the following documents:

1. IRB Application Format
2. Cvs of Drs. Alok S, Mohankumar, Srujan.
3. No. of Documents 1- 2

The following Institutional Review Board (**Silver**, Research & Ethics Committee) members were present at the meeting held on October 30, 2019 at 9.45 am in the New IRB Room, Christian Medical College, Bagayam, Vellore -632002.

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

Name	Qualification	Designation	Affiliation
Dr. George Thomas	D Ortho, PhD	Orthopaedic Surgeon, St. Isabella Hospital, Chennai, Chairperson, Ethics Committee, IRB, Chennai	External, Clinician
Prof. Keith Gomez	BSc, MA (S.W), M. Phil (Psychiatry Social Work)	Student counselor, Loyola College, Chennai, Deputy Chairperson, Ethics Committee, IRB	External, Lay Person & Social Scientist
Dr. Biju George	MBBS, MD, DM	Professor, Haematology, Additional Vice Principal (Research), Deputy Chairperson (Research Committee), Member Secretary (Ethics Committee), IRB, CMC, Vellore.	Internal, Clinician
Dr. P. Zachariah	MBBS, PhD	Retired Professor, Vellore	External, Clinician
Dr. Sathya Subramani	MD, PhD	Professor, Physiology, CMC, Vellore	Internal, Clinician
Dr. Prasanna Samuel	M. Sc, PhD	Lecturer, Biostatistics, CMC, Vellore	Internal, Statistician
Dr. Jayaprakash Muliyil	MBBS, MD, MPH, Dr PH (Epid), DMHC	Retired Professor, Vellore	External, Scientist & Epidemiologist
Rev. Dr. T. Arul Dhas	MSc, BD, DPC, PhD(Edin)	Chaplaincy Department, CMC, Vellore	Internal, Social Scientist
Dr. Blessed Winston	MBBS., MD	Associate Professor, Clinical Pharmacology, CMC, Vellore	Internal, Pharmacologist
Dr. Suresh Devasahayam	BE, MS, PhD	Professor of Bio-Engineering, CMC, Vellore	Internal, Basic Medical Scientist
Dr. P. Zachariah	MBBS, PhD	Retired Professor, Vellore	External, Clinician
Dr. Ajith Sivadasan	MD, DM	Professor, Neurological Sciences, CMC, Vellore	Internal, Clinician
Mr. Samuel Abraham	MA, PGDBA, PGDPM, M. Phil, BL.	Sr. Legal Officer, Vellore	External Legal Expert
Mr. C. Sampath	BSc, BL	Advocate, Vellore	External, Legal Expert
Mrs. Ilavarasi Jesudoss	M Sc (Nursing)	Deputy Nursing Superintendent, College of Nursing, CMC, Vellore	Internal, Nurse
Mrs. Pattabiraman	BSc, DSSA	Social Worker, Vellore	External, Lay person

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

Dr. Abhay Gahukamble	MS, D Ortho, DNB(Ortho)	Associate Professor, Paediatric Orthopaedics, CMC, Vellore	Internal, Clinician
Dr. Suceena Alexander	MBBS, MD, DM	Associate Professor, Nephrology, CMC, Vellore	Internal, Clinician
Dr. Sujith J Chandy	MD., PhD., FRCP (E)	Professor, Clinical Pharmacology, CMC, Vellore	Internal, Pharmacologist
Dr. Shirley David	MSc, PhD	Professor, Head of Fundamentals Nursing Department, College of Nursing, CMC, Vellore	Internal, Nurse

We approve the project to be conducted as presented.

Kindly provide the total number of patients enrolled in your study and the total number of Withdrawals for the study entitled: "Recreating HPFH-associated point mutation to elevate the fetal haemoglobin levels for the treatment of beta-hemoglobinopathies using CRISPR based tools"on a monthly basis. Please send copies of this to the Research Office (research@cmcvellore.ac.in).

The Institutional Ethics Committee expects to be informed about the progress of the project, any **adverse events** occurring in the course of the project, any **amendments in the protocol and the patient information / informed consent**. On completion of the study you are expected to submit a copy of the **final report**. Respective forms can be downloaded from the following link: http://172.16.11.136/Research/IRB_Polices.html in the CMC Intranet and in the CMC website link address: http://www.cmch-vellore.edu/static/research/Index.html.

Administrative Committee's approval is to be obtained for opening the account-head, employing any personnel or purchasing any equipment. The investigator also needs to present to Administrative Committee, the terms and condition of the Funding agency for approval

Yours sincerely,

Dr. Biju George
Secretary (Ethics Committee)
Institutional Review Board.

Dr. BIJU GEORGE
MBBS., MD., DM.
SECRETARY - (ETHICS COMMITTEE)
Institutional Review Board,
Christian Medical College, Vellore - 632 002.

IRB Min. No. 12309 (OTHER) dated 30.10.2019

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

OFFICE OF THE VICE PRINCIPAL (RESEARCH)
INSTITUTIONAL BIO-SAFETY COMMITTEE (IBSC)
CHRISTIAN MEDICAL COLLEGE, VELLORE - 632 002, INDIA

22nd September, 2020

To,

Dr. Mohankumar K. Murugesan,
Assistant Investigator,
Centre for Stemcell Research,
CMC

Dear Dr. Mohankumar K. Murugesan,

The Institutional Bio Safety Committee of the Christian Medical College, Vellore, reviewed and discussed your project entitled *"Recreating HPFH-associated point mutation to elevate the fetal haemoglobin levels for the treatment of beta-hemoglobinopathies using CRISPR based tools."*

IRB Min No: 12309 dated 30.10.2019

The following members were present at the meeting on organised on 12th September, 2020 by 10.00 am in the IRB Meeting Room, College Campus, Christian Medical College, Vellore 632 004.

S.No	Name	Designation	IBSC Affiliation
1	Dr. Anna B Pulimood	Professor & Principal, CMC.	Chairperson
2	Dr. Suceena Alexander	Professor & Addl. Vice Principal (Research), CMC.	Member Secretary
3	Dr. Rayala Suresh Kumar	Professor, Department of Biotechnology, IIT, CMC.	DBT Nominee
4	Dr. Molly Jacob	Professor & Head, Department of Biochemistry, CMC.	Internal Expert
5	Dr. Alok Srivastava	Professor & Head, CSCR, CMC.	Internal Expert
6	Dr. Joy S Michael	Professor, Microbiology, CMC.	Biosafety Officer
7	Dr. R V Shaji	Professor and Adjunct Scientist, CSCR, CMC	Internal Expert

8	Dr. Sitara Swarna Rao A	Professor, Wellcome trust labs, CMC	Internal Expert
9	Dr. K A Balasubramanian	Emeritus Professor, Sri Narayani Hospital and Research Centre	External Expert

As there are no biosafety concerns, the IBSC approval has been given for the study.

Yours sincerely,

SECRETARY
Institutional Bio-Safety Committee
Christian Medical College,
Vellore - 632 002, Tamil Nadu, India.

Dr. Suceena Alexander
Secretary, Institutional Bio Safety Committee
Christian Medical College, Vellore

INSTITUTIONAL ANIMAL ETHICS COMMITTEE
CHRISTIAN MEDICAL COLLEGE, VELLORE

Dr. Anna B. Pulimood
Chairperson

Dr. Joe Varghese
Member Secretary

Date: 20.06.2020

Dr. Mohankumar,
Assistant Investigator,
Centre for Stem Cell Research,
CMC, Vellore.

Sub: Proposal for IAEC approval (NEW) – "***Recreating HPFH-associated point mutation to evaluate the fetal haemoglobin levels for the treatment of beta-hemoglobinopathies using CRISPR-based tools***"

Dear Dr. Mohankumar,

Your application to the Institutional Animal Ethics Committee (IAEC) titled "*Recreating HPFH-associated point mutation to evaluate the fetal haemoglobin levels for the treatment of beta-hemoglobinopathies using CRISPR based tools*" has been reviewed by the IAEC at the meeting held on 20/06/2020.

The following members of the IAEC were present at the meeting:

1. Dr. A. Yasotha, CPCSEA Main Nominee
2. Dr. D. Sivaraman, Scientist from outside the institute
3. Dr. B.R. Senthilkumar, Socially aware member
4. Dr. Christhunesa Christudass, Scientist
5. Dr. Suceena Alexander, Scientist
6. Dr. K. Imayarasi, Veterinarian
7. Dr. Joe Varghese, Member Secretary

After discussion, the project proposal was approved. Thirty seven (37) NBSGW mice have been sanctioned for Year 1 of the study.

The IAEC approval number for the study is **10/2020**

Please see overleaf for responsibilities of the principal investigator.

With best wishes,

Yours sincerely,

Dr. Anna B. Pulimood
Chairperson

Dr. Joe Varghese
Member Secretary

SECRETARY
Institutional Animal Ethics Committee,
Christian Medical College,
Bagayam, Vellore - 632 002, Tamil Nadu.

INSTITUTIONAL ANIMAL ETHICS COMMITTEE
CHRISTIAN MEDICAL COLLEGE, VELLORE

Dr. Anna B. Pulimood
Chairperson

Dr. Joe Varghese
Member Secretary

Responsibilities of the Principal Investigator:

1. Progress Report / Project completion report should be submitted for consideration of IAEC at prescribed intervals for review and it should not exceed the timeline of the research as mentioned in the Form B / IAEC approval certificate
2. Final report should be submitted to the IAEC of the establishment at the end of study.
3. All Serious Adverse Events (SAEs) and the interventions undertaken should be intimated to IAEC.
4. Protocol deviation, if any, should be informed with adequate justifications to IAEC and in case of large animals, it should be intimated to CPCSEA immediately for consideration of CPCSEA.
5. The procedural deviations in research protocol shall be treated as a fresh research protocol by CPCSEA.
6. Any new information related to the study should be communicated to IAEC immediately.
7. Premature termination of study should be notified to the IAEC with reasons along with summary of the data obtained so far.
8. Change of investigators / sites should be informed and approval of IAEC should be undertaken first.

Dr. Anna B. Pulimood
Chairperson

Dr. Joe Varghese
Member Secretary

SECRETARY
Institutional Animal Ethics Committee,
Christian Medical College,
Bagayam, Vellore - 632 002, Tamil Nadu.

INSTITUTIONAL ANIMAL ETHICS COMMITTEE
CHRISTIAN MEDICAL COLLEGE, VELLORE

Dr. Anna B. Pulimood
Chairperson

Dr. Joe Varghese
Member Secretary

Date: 19.06.2021

Dr. Mohankumar K Murugesan
Assistant Investigator
Centre for Stem Cell Research
CMC, Vellore.

Sub: Project Progress report - "***Recreating HPFH-associated point mutation to elevate the fetal haemoglobin levels for the treatment of beta hemoglobinopathies using CRISPR based tools***."

Ref: IAEC no. 1**0/2020**

Dear Dr.Mohankumar,

The progress report submitted by you to the Institutional Animal Ethics Committee (IAEC) for the project titled "***Recreating HPFH-associated point mutation to elevate the fetal haemoglobin levels for the treatment of beta hemoglobinopathies using CRISPR based tools***" has been reviewed by the IAEC at the meeting held on 19/06/2021.

The following members of the IAEC were present at the meeting:

1. Dr. Anna B. Pulimood, Chairperson, IAEC (Internal member)
2. Dr. A. Yasotha, CPCSEA Main Nominee (External member)
3. Dr. D. Sivaraman, Scientist from outside the institute (External member)
4. Dr. B.R. Senthilkumar, Socially aware member (External member)
5. Dr. Christhunesa Christudass, Scientist (Internal member)
6. Dr. K. Imayarasi, Veterinarian (Internal member)
7. Dr. Joe Varghese, Member Secretary (Internal member)

After discussion, seventy-four (74) NBSGW mice have been sanctioned for Year 2 of the study.

Please see overleaf for responsibilities of the principal investigator.

With best wishes,

Yours sincerely,

Dr. A. Yasotha
CPCSEA Main Nominee

Dr. Anna B. Pulimood
Chairperson

CHAIRPERSON
Institutional Animal Ethics Committee,
Christian Medical College,
Bagayam, Vellore - 632 002, Tamil Nadu.

Dr. Joe Varghese
Member Secretary

SECRETARY
Institutional Animal Ethics Committee,
Christian Medical College,
Bagayam, Vellore - 632 002, Tamil Nadu.

INSTITUTIONAL ANIMAL ETHICS COMMITTEE
CHRISTIAN MEDICAL COLLEGE, VELLORE

Dr. Solomon Sathishkumar
Chairperson

Dr. Joe Varghese
Member Secretary

Date: 16.07.2022

Dr. Mohankumar K,
Assistant Investigator,
Centre for Stem Cell Research,
CMC, Vellore.

Sub: Project Progress report - "Recreating HPFH associated point mutation to elevate the fatal hemoglobin levels for the treatment of beta hemoglobinopathies using CRISPR tools."

Ref: IAEC no. **10/2020**

Dear Dr. Mohan,

The progress report submitted by you to the Institutional Animal Ethics Committee (IAEC) titled "***Recreating HPFH associated point mutation to elevate the fatal hemoglobin levels for the treatment of beta hemoglobinopathies using CRISPR tools***" has been reviewed by the IAEC at the meeting held on 16/07/2022.

The following members of the IAEC were present at the meeting:

1. Dr. P.Kumarasamy, CPCSEA Main Nominee (External Member)
2. Dr. P.Patric Joshua, Scientist from outside the institute (External Member)
3. Dr. R.Prakash, Socially aware member (External Member)
4. Dr.Christhunesa Soundararajan C, Scientist from a different discipline (Internal Member)
5. Dr. Eunice Sindhuvi, Scientist from a different biological discipline (Internal Member)
6. Dr. K. Imayarasi, Veterinarian (Internal Member)
7. Dr. Joe Varghese, Member Secretary IAEC, CMC, Vellore (Internal Member)

After discussion, sixty (60) NBSGW mice have been approved for Second Year of the study.

Please see overleaf for responsibilities of the principal investigator.

With best wishes,

Yours sincerely,

Dr. Solomon Sathishkumar
Chairperson

CHAIRPERSON
Institutional Animal Ethics Committee,
Christian Medical College,
Bagayam, Vellore - 632 002, Tamil Nadu.

Dr. Joe Varghese
Member Secretary

SECRETARY
Institutional Animal Ethics Committee,
Christian Medical College,
Bagayam, Vellore - 632 002, Tamil Nadu.

INSTITUTIONAL ANIMAL ETHICS COMMITTEE
CHRISTIAN MEDICAL COLLEGE, VELLORE

Dr. Solomon Sathishkumar
Chairperson

Dr. Joe Varghese
Member Secretary

Responsibilities of the Principal Investigator:

1. Progress Report / Project completion report should be submitted for consideration of IAEC at prescribed intervals for review and it should not exceed the timeline of the research as mentioned in the Form B / IAEC approval certificate.
2. Final report should be submitted to the IAEC of the establishment at the end of the study.
3. All Serious Adverse Events (SAEs) and the interventions undertaken should be intimated to IAEC.
4. Protocol deviation, if any, should be informed with adequate justifications to IAEC and in the case of large animals, it should be intimated to CPCSEA immediately for consideration of CPCSEA.
5. The procedural deviations in the research protocol shall be treated as a fresh research protocol by CPCSEA.
6. Any new information related to the study should be communicated to IAEC immediately.
7. Premature termination of study should be notified to the IAEC with reasons along with summary of the data obtained so far.
8. Change of investigators / sites should be informed and approval of IAEC should be undertaken first.

Dr. Solomon Sathishkumar
Chairperson

Dr. Joe Varghese
Member Secretary

OFFICE OF RESEARCH
INSTITUTIONAL REVIEW BOARD (IRB)
CHRISTIAN MEDICAL COLLEGE, VELLORE, INDIA

Ethics Committee Registration No: ECR/326/INST/TN/2013 Re Reg-2016 Issued under Rule 122D of the Drugs & Cosmetics Rules 1945, Govt. of India

Dr. George Thomas, M.B.B.S., D. Ortho., Ph.D.,
Chairperson, Ethics Committee

Dr. L. Jeyaseelan, M.Sc., Ph.D., FSMS, FRSS.,
Secretary, Research Committee

Prof. Keith Gomez, B.Sc., MA (S.W), M.Phil.,
Deputy Chairperson, Ethics Committee

Dr. Anna Benjamin Pulimood, M.B.B.S., MD., Ph.D.,
Chairperson, Research Committee & Principal

Dr. Biju George, M.B.B.S., MD., DM.,
Deputy Chairperson,
Secretary, Ethics Committee, IRB
Additional Vice-Principal (Research)

May 14, 2019

Dr. Mohankumar K. Murugesan,
Assistant Investigator,
Department of CSCR,
Christian Medical College,
Vellore – 632 004

Sub: **External Research Grant: ICMR**
Genome engineering hematopoietic stem cells for the treatment of Hemophilia A
Dr. Mohankumar K. Murugesan , Assistant Investigator , Centre for Stem Cell Research , Dr. Sanjay Kumar, Scientist, CSCR , Dr. Saravanabhavan Thangavel, Assistant Investigator, CSCR Dr. Alok Srivastava, Hematology & CSCR

Ref: IRB Min. No. 11943 (OTHER) dated 27.03.2019

Dear Dr. Mohankumar K. Murugesan,

The Institutional Review Board (**Silver**, Research and Ethics Committee) of the Christian Medical College, Vellore, reviewed and discussed your project titled "Genome engineering hematopoietic stem cells for the treatment of Hemophilia A" on March 27, 2019. I am quoting below the minutes of the meeting:

The Committee reviewed the following documents:

1. IRB Application format
2. Patient information sheet and informed Consent form
3. Cvs of . Drs. Alok Srivastava, Mohan Kumar Sanjay and Saravanabhavan.
4. No. of documents 1 – 3

The following Institutional Review Board (**Silver**, Research & Ethics Committee) members were present at the meeting held on March 27, 2019 at 9.45 am in the New IRB Room, Christian Medical College, Bagayam, Vellore -632002.

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

Name	Qualification	Designation	Affiliation
Dr. George Thomas	MBBS, D Ortho, PhD	Orthopaedic Surgeon, St. Isabella Hospital, Chennai, Chairperson, Ethics Committee, IRB, Chennai	External, Clinician
Prof. Keith Gomez	BSc, MA (S.W), M. Phil (Psychiatry Social Work)	Student counselor, Loyola College, Chennai, Deputy Chairperson, Ethics Committee, IRB	External, Lay Person & Social Scientist
Dr. Biju George	MBBS, MD, DM	Professor, Haematology, Additional Vice Principal (Research), Deputy Chairperson (Research Committee), Member Secretary (Ethics Committee), IRB, CMC, Vellore.	Internal, Clinician
Dr. L. Jeyaseelan	MSc, PhD, FSMS, FRSS	Professor & Head, Biostatistics, Secretary (Research Committee), IRB, CMC, Vellore	Internal, Statistician
Dr. Sujith J Chandy	MBBS., MD., PhD., FRCP (E)	Professor, Clinical Pharmacology, CMC, Vellore	Internal, Pharmacologist
Rev. Dr. T. Arul Dhas	MSc, BD, DPC, PhD(Edin)	Chaplaincy Department, CMC, Vellore	Internal, Social Scientist
Dr. Suresh Devasahayam	BE, MS, PhD	Professor of Bio-Engineering, CMC, Vellore	Internal, Basic Medical Scientist
Dr. P. Zachariah	MBBS, PhD	Retired Professor, Vellore	External, Clinician
Dr. D. J. Christopher	B. Sc, MBBS, DTCD DNB, FRCP(Glasg), FCCP(USA)	Professor, Pulmonary Medicine, CMC, Vellore	Internal, Clinician
Dr. RV. Shaji	B.Sc, M.Sc, PhD	Professor, Heamatology, CMC, Vellore	Internal, Basic Medical Scientist
Dr. Niranjan Thomas	MBBS, MD (Paed) DCH, DNB	Professor, Paediatrics, CMC, Vellore	Internal, Clinician
Dr. Ashish Goel	MBBS, MD, DM	Professor, Hepatology, CMC, Vellore	Internal, Clinician
Mr. C. Sampath	BSc, BL	Advocate, Vellore	External, Legal Expert

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

Mrs. Pattabiraman	BSc, DSSA	Social Worker, Vellore	External, Lay person
Dr. Prasanna Samuel	MSc, PhD	Lecturer, Biostatistics, CMC, Vellore	Internal, Statistician
Dr. Abhay Gahukamble	MS, D Ortho, DNB(Ortho)	Associate Professor, Paediatric Orthopaedics, CMC, Vellore	Internal, Clinician
Dr. Sridhar Gibikote	MBBS, DMRD, DNB	Professor, Radiology, CMC, Vellore	Internal, Clinician
Dr. Suceena Alexander	MBBS, MD, DM	Associate Professor, Nephrology, CMC, Vellore	Internal, Clinician

We approve the project to be conducted as presented.

Kindly provide the total number of patients enrolled in your study and the total number of Withdrawals for the study entitled: "Genome engineering hematopoietic stem cells for the treatment of Hemophilia A" on a monthly basis. Please send copies of this to the Research Office (research@cmcvellore.ac.in).

Administrative Committee's approval is to be obtained for opening the account-head, employing any personnel or purchasing any equipment. The investigator also needs to present to Administrative Committee, the terms and condition of the Funding agency for approval

Yours sincerely,

Dr. Biju George
Secretary (Ethics Committee)
Institutional Review Board.

Dr. BIJU GEORGE
MBBS., MD., DM.
SECRETARY - (ETHICS COMMITTEE)
Institutional Review Board,
Christian Medical College, Vellore - 632 002.

IRB Min. No. 11943 (OTHER) dated 27.03.2019

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

OFFICE OF RESEARCH
INSTITUTIONAL REVIEW BOARD (IRB)
CHRISTIAN MEDICAL COLLEGE, VELLORE, INDIA
CDSCO - Ethics Committee Registration No: ECR/326/INST/TN/2013/RR-2019, DHR Registration No: EC/NEW/INST/2023/TN/0211

Dr. J. Amalorpavanathan, M.S, Dip. NBE, M. Ch.,
Chairperson, Ethics Committee

Dr. Prasanna Samuel, M.Sc., Ph.D.,
Secretary, Research Committee

Prof. Keith Gomez, MA (S.W), M.Phil.,
Deputy Chairperson, Ethics Committee.

Dr. Jacob John, MD., Ph D.,
Chairperson, Research Committee

Dr. Suceena Alexander, MD, DM (Nephrology), FRCP (Lon), FASN., Ph.D.
Secretary, Ethics Committee, IRB
Additional Vice-Principal (Research)

Ref: IRB – A16 – 28.06.2023

September 14, 2023

Dr. Mohankumar K. Murugesan,
Assistant Investigator,
Centre for Stem Cell Research,

Christian Medical College,
Vellore.

Ref: IRB Min. No. 11943 (other) dated 27/03/2019

Dear Dr. Mohankumar K. Murugesan,

The Institutional Review Board (Silver, Research and Ethics Committee) of the Christian Medical College, Vellore, reviewed and discussed for the study titled "Genome engineering hematopoietic stem cells for the treatment of Hemophilia A" on June 28, 2023.

(a Title: (page number: 1, also throughout the study FVIII is changed to FIX (changed on 53-21))
Genome engineering hematopoietic stem cells for the treatment of Hemophilia B (reason stated below)

(b Protocol: (page no:7,8 and 9 (changed on 5-3-21))

- We like to use Adeno Associated Virus (AAV) as a donor template delivery system instead of episomal plasmid via electroporation.
- The important reason for this improvisation is the toxicity caused by episomal plasmid delivery via nucleofection.
- In case of AAV donor the toxicity is reduced and the efficiency of HDR is greatly increased, so many research communities in genome editing field have switched to this protocol [1-7].
- One major disadvantage with AAV its packaging limit. It can only pack upto 4.7kb of genetic material of our interest which needs to be knocked in.
- In our IRB we have attempted to correct hemophilia A (Factor VIII) via homologous directed repair (HDR) in ITGA2B locus (under platelet promoter). • The factor VIII CDS is around 4.4kb, along with homology arm it will be 5.4kb which is above the packaging limit of AAV.
- To overcome this problem, we thought of knocking in Factor IX (Hemophilia B) due to its smaller CDS (1.4 KB), along with homology arm it will be around 2.4 kb which is within the limit of AVV packaging.
- References:

1. Rai, R. et al. Targeted gene correction of human hematopoietic stem cells for the treatment of Wiskott - Aldrich Syndrome. Nat. Commun. 11, (2020).
2. Pavani, G. et al. Ex vivo editing of human hematopoietic stem cells for erythroid expression of therapeutic proteins. Nat. Commun. 11, (2020).
3. De Ravin, S. S. et al. CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-Iinked chronic granulomatous disease. Sci. Transl. Med. 9, (2017).
4. Dever, D. P. et al. CRISPR/Cas9 ß-globin gene targeting in human haematopoietic stem cells. Nature 539, 384-389 (2016).
5. Bak, R. O. et al. Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. Elife 6, 1—19 (2017).
6. Wang, J. et al. Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. Nat. Biotechnol. 33, 1256— 1263 (2015).
7. Hiramoto, T. , Li, L. B. , Funk, S. E. , Hirata, R. K. & Russell, D. W. Nuclease-free Adeno-Associated Virus-Mediated 112rg Gene Editing in X-SCID Mice. Mol. Ther. 26, 1255-1265 (2018).

7. Will any of these amendments significantly alter the balance of risks and benefits to study participants? (please describe)
   - Since the genome editing will be invitro and then edited cells will be delivered to the participants (Ex-vivo delivery) it will not cause any immune response.
   - Also, the AAV does not integrate into the host genome, there is no risk involving insertional mutagenesis.
   - One of the main benefits compared to the previous protocol is that the viability of the edited cells and the editing efficiency will be increase drastically, which will improve the desired effect and the engraftment potential.

The following Institutional Review Board (Silver, Research & Ethics Committee) members were present at the meeting held on June 28, 2023 at 09.45 am in the New IRB Room, Christian Medical College, Bagayam, Vellore-632002.

ETHICS COMMITTEE MEMBERS			
Name	**Qualification**	**Designation**	**Affiliation**
Dr. George Thomas	D. Ortho, Ph.D.	Vice Chairperson, Ethics Committee, IRB, Orthopedic Surgeon, CMC, Ranipet Campus,	Internal, Clinician

Dr. J. Amalorpavanathan, M.S. Dip. NBE, M. Ch..
Chairperson, Ethics Committee

Dr. Prasanna Samuel, M.Sc., Ph.D..
Secretary, Research Committee

Prof. Keith Gomez, MA (S.W), M.Phil..
Deputy Chairperson, Ethics Committee.

Dr. Jacob John, MD., Ph D.,
Chairperson, Research Committee

Dr. Suceena Alexander, MD, DM (Nephrology), FRCP (Lon), FASN., Ph.D.
Secretary, Ethics Committee, IRB
Additional Vice-Principal (Research)

Dr. Suceena Alexander	M.D, D.M (Nephro), FRCP (Lon.), FASN, Ph.D.	Secretary – (Ethics Committee), IRB, Addl. Vice Principal (Research), Professor of Nephrology, CMC, Vellore	Internal, Clinician
Prof. Keith Gomez	M.A (S.W), M. Phil (Psychiatry Social Work)	Deputy Chairperson, Ethics Committee, IRB, Student counselor, Loyola College, Chennai.	External, Social Scientist
Dr. Jayaprakash Muliyil	M.D, MPH, Dr. PH (Epid), DMHC	Retired Professor, CMC, Vellore	External, Scientist & Epidemiologist
Dr. Blessed Winston	M.D Pharmacology	Associate Professor, Clinical Pharmacology, CMC, Vellore	Internal, Pharmacologist
Mrs. Anne Jarone	M. Sc (Nursing)	Professor, Associate Nursing Superintendent, CMC, Vellore	Internal, Nurse
Mrs. Mary Anbarasi Johnsons	M. Sc., Ph.D	Professor & Head, Paediatric Nursing College of Nursing, CMC, Vellore	Internal, Nurse
Rev. Joseph Devaraj	B. Sc, BD	Chaplaincy Department, CMC, Vellore	Internal, Social Scientist
Mr. C. Sampath	B.Sc, BL	Sr. Legal Officer, Vellore	External, Legal Expert
Mrs. B. Scholastica Mary Vithiya	M. Phil, Ph. D.	Assistant Professor, Auxilium College, Vellore	External, Layperson
RESEARCH COMMITTEE MEMBERS			
Dr. D. J. Christopher	DTCD DNB, FRCP(Glasg), FCCP(USA)	Professor, Pulmonary Medicine, CMC, Vellore	Internal, Clinician

Dr. Jacob John	MD, MPH	Chairperson, Research Committee, Professor, Community Medicine, CMC, Vellore	Internal, Clinician
Dr. Prasanna Samuel	M. Sc, Ph.D.	Secretary, Research Committee, Associate Professor of Biostatistics, CMC, Vellore	Internal, Statistician
Dr. Rajdeep Ojha	M. Tech, PhD	Associate Professor of Physical Medicine and Rehabilitation, CMC, Vellore	Internal, Basic Medical Scientist
Dr. Sridhar Gibikote	DMRD, DNB	Professor, Radiology, CMC, Vellore	Internal, Clinician
Dr. Joe Varghese	MD, Ph.D.,	Professor, Biochemistry, CMC	Internal, Basic Medical Scientist
Dr. Santhanam Sridhar	DCH, DNB	Professor, Neonatology, CMC, Vellore	Internal, Clinician
Dr. Nihal Thomas	MD MNAMS DNB (Endo) FRACP (Endo) FRCP (Edin) FRCP (Glas) FRCP (London) FACP Ph.D. (Copenhagen)	Professor & Head Department of Endocrinology, Diabetes, and Metabolism	Internal, Clinician
Dr. Christhunesa S. Christudass	M.Sc., Ph.D	Professor, Neurochemistry, Department of Neurological Sciences	Internal, Basic Medical Scientist
Dr. Rohin Mittal	MS, DNB	Professor, Department of General Surgery, CMC Vellore	Internal Clinician
Dr. Elizabeth Vinod	MBBS, MD,	Associate Professor, Department of Physiology, CMC Vellore	Internal Basic Medical Scientist

We approve the above amendment as presented.

Yours sincerely,

Dr. Suceena Alexander
Secretary (Ethics Committee)
Institutional Review Board.

Dr. SUCEENA ALEXANDER
MD, DM(Nephrology), FRCP(Lon), FASN, PhD.
Secretary - (Ethics Committee),
Institutional Review Board, Christian Medical College,
Vellore - 632 002, Tamil Nadu, India.

IRB Min. No. 11943 (other) dated 27/03/2019

Ethics Committee Silver, Office of Research, Ist Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284289, 3075645 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

OFFICE OF THE VICE PRINCIPAL (RESEARCH)
CHRISTIAN MEDICAL COLLEGE, VELLORE - 632 004, INDIA

Date: 26th October, 2019

To,

Dr. Mohankumar Murugesan,
Assist. Investigator,
Centre for Stem Cell Research,
CMC

Dear Dr. Mohankumar Murugesan,,

The Institutional Bio Safety Committee of the Christian Medical College, Vellore, reviewed and discussed your project entitled "Genome engineering hematopoietic stem cells for the treatment of Hemophilia A"

IRB Min. No: 11943 dated 27.03.2019

The following members were present at the meeting on organised on 23rd October, 2019 by 1.00 am in the IRB Meeting Room, College Campus, Christian Medical College, Vellore 632 004.

S. No	Name	Designation	IBSC Affiliation
1	Dr. Anna B Pulimood	Professor & Principal, CMC.	Chairperson
2	Dr. Biju George	Professor & Addl. Vice Principal (Research), CMC.	Member Secretary
3	Dr. Rayala Suresh Kumar	Professor, Department of Biotechnology, IIT, CMC.	DBT Nominee
4	Dr. K A Balasubramanian	Professor (Retd.).	External Expert
5	Dr. Alok Srivastava	Professor & Head, CSCR	Internal Expert
6	Dr. Molly Jacob	Professor & Head, Department of Biochemistry, CMC.	Internal Expert

7	Dr. Rajesh Kannangai	Professor & Head, Department of Virology, CMC.	Internal Expert
8	Dr. Sitara Swarna Rao A	Professor, Wellcome Trust labs, CMC.	Internal Expert

After discussion, it was resolved to **ACCEPT the proposal. There is no Biosafety hazard in this study.**

Yours sincerely,

Dr. BIJU GEORGE
MBBS., MD., DM.
Vice - Principal (Research)
Christian Medical College, Vellore - 632 002.

Dr. Biju George
Secretary, Institutional Bio Safety Committee
Christian Medical College, Vellore

OFFICE OF RESEARCH
INSTITUTIONAL REVIEW BOARD (IRB)
CHRISTIAN MEDICAL COLLEGE, VELLORE, INDIA

Ethics Committee Registration No: ECR/326/INST/TN/2013 Re Reg-2016 Issued under Rule 122D of the Drugs & Cosmetics Rules 1945, Govt. of India

Dr. George Thomas, D. Ortho., Ph.D.,
Chairperson, Ethics Committee

Dr. Antonisamy, Ph.D., FSMS, FRSS.,
Secretary, Research Committee

Prof. Keith Gomez, MA (S.W), M.Phil.,
Deputy Chairperson, Ethics Committee

Dr. Anna Benjamin Pulimood, MD., Ph.D.,
Chairperson, Research Committee & Principal

Dr. Biju George, MD., DM.,
Secretary, Ethics Committee, IRB
Additional Vice-Principal (Research)

May 27, 2019

Dr. Mohankumar K. Murugesan,
Assistant Investigator,
Department of CSCR,
Christian Medical College,
Vellore – 632 004

Sub: **External Research Grant: ICMR**
Genome engineering hematopoietic stem cells for the treatment of Hemophilia A
Dr. Mohankumar K. Murugesan , Assistant Investigator, Centre for Stem Cell Research , Dr. Sanjay Kumar, Scientist, CSCR , Dr. Saravanabhavan Thangavel, Assistant Investigator, CSCR Dr. Alok Srivastava, Hematology & CSCR

Ref: IRB Min. No. 11943 (OTHER) dated 27.03.2019

Dear Dr. Mohankumar K. Murugesan,

Institutional Committee for Stem Cell Research and Therapy (IC-SCRT/RED IRB) of the Christian Medical College, Vellore, reviewed and discussed your project titled "Genome engineering hematopoietic stem cells for the treatment of Hemophilia A" on March 27, 2019. I am quoting below the minutes of the meeting.

The Committee reviewed the following documents:

1. IRB Application format
2. Patient information sheet and informed Consent form
3. Cvs of. Drs. Alok Srivastava, Mohan Kumar Sanjay and Saravanabhavan.
4. No. of documents 1 – 3

Institutional Committee for Stem Cell Research and Therapy (IC-SCRT/RED IRB) members were present at the meeting held on March 27, 2019 at 9.45 am in the New IRB Room, Christian Medical College, Bagayam, Vellore -632002.

Ethics Committee Silver, Office of Research, I Floor, Carman Block, Christian Medical College, Vellore, Tamil Nadu 632 002
Tel: 0416 – 2284294, 2284202 Fax: 0416 – 2262788 E-mail: research@cmcvellore.ac.in

Name	Qualification	Designation	Affiliation
Dr. George Thomas	MBBS, D Ortho, PhD	Orthopaedic Surgeon, St. Isabella Hospital, Chennai, Chairperson, Ethics Committee, IRB, Chennai	External, Clinician
Prof. Keith Gomez	BSc, MA (S.W), M. Phil (Psychiatry Social Work)	Student counselor, Loyola College, Chennai, Deputy Chairperson, Ethics Committee, IRB	External, Lay Person & Social Scientist
Dr. Biju George	MBBS, MD, DM	Professor, Haematology, Additional Vice Principal (Research), Deputy Chairperson (Research Committee), Member Secretary (Ethics Committee), IRB, CMC, Vellore.	Internal, Clinician
Dr. Suresh Rayala	M. Sc, Ph.D (Biochemistry)	IIT Madras, Chennai	External Stem Cell Expert
Dr. Sujith J Chandy	MBBS., MD., PhD., FRCP (E)	Professor, Clinical Pharmacology, CMC, Vellore	Internal, Pharmacologist
Rev. Dr. T. Arul Dhas	MSc, BD, DPC, PhD(Edin)	Chaplaincy Department, CMC, Vellore	Internal, Social Scientist
Dr. RV. Shaji	B.Sc, M.Sc, PhD	Professor, Heamatology, CMC, Vellore	Internal, Basic Medical Scientist
Mr. C. Sampath	BSc, BL	Advocate, Vellore	External, Legal Expert
Mrs. Pattabiraman	BSc, DSSA	Social Worker, Vellore	External, Lay person

We approve the project to be conducted as presented.

Kindly provide the total number of patients enrolled in your study and the total number of Withdrawals for the study entitled: "Genome engineering hematopoietic stem cells for the treatment of Hemophilia A" on a monthly basis. Please send copies of this to the Research Office (research@cmcvellore.ac.in).

Yours sincerely,

Dr. BIJU GEORGE
MBBS., MD., DM.
SECRETARY - (ETHICS COMMITTEE)
Institutional Review Board,
Medical College, Vellore - 632 002.

Dr. Biju George
Secretary (Ethics Committee)
Institutional Review Board.

IRB Min. No. 11943 (OTHER) dated 27.03.2019

Appendix b - Publications

RESEARCH ARTICLE

Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin

Nithin Sam Ravi[1,2], Beeke Wienert[3,4,5], Stacia K Wyman[3], Henry William Bell[5], Anila George[1,2], Gokulnath Mahalingam[1], Jonathan T Vu[3], Kirti Prasad[1,6], Bhanu Prasad Bandlamudi[1], Nivedhitha Devaraju[1,6], Vignesh Rajendiran[1,2], Nazar Syedbasha[1], Aswin Anand Pai[2,7], Yukio Nakamura[8], Ryo Kurita[9], Muthuraman Narayanasamy[1,10], Poonkuzhali Balasubramanian[2,7], Saravanabhavan Thangavel[1], Srujan Marepally[1], Shaji R Velayudhan[1,2,7], Alok Srivastava[1,2,7], Mark A DeWitt[3,11], Merlin Crossley[5], Jacob E Corn[3,12], Kumarasamypet M Mohankumar[1,2]*

[1]Centre for Stem Cell Research (a Unit of inStem, Bengaluru), Christian Medical College Campus, Vellore, India; [2]Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, India; [3]Innovative Genomics Institute, University of California, Berkeley, Berkeley, United States; [4]Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, United States; [5]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia; [6]Manipal Academy of Higher Education, Karnataka, India; [7]Department of Haematology, Christian Medical College & Hospital, Vellore, India; [8]Cell Engineering Division, RIKEN BioResource Center, Ibaraki, Japan; [9]Research and Development Department, Central Blood Institute Blood Service Headquarters, Japanese Red Cross Society, Japan, Tokyo, Japan; [10]Department of Biochemistry, Christian Medical College, Vellore, India; [11]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, United States; [12]Institute of Molecular Health Sciences, Department of Biology, Zurich, Switzerland

***For correspondence:** mohankumarkm@cmcvellore.ac.in

Competing interest: The authors declare that no competing interests exist.

Funding: See page 22

Preprinted: 01 July 2020
Received: 03 December 2020
Accepted: 11 February 2022
Published: 11 February 2022

Reviewing Editor: Stephen C Ekker, Mayo Clinic, United States

Abstract Naturally occurring point mutations in the *HBG* promoter switch hemoglobin synthesis from defective adult beta-globin to fetal gamma-globin in sickle cell patients with hereditary persistence of fetal hemoglobin (HPFH) and ameliorate the clinical severity. Inspired by this natural phenomenon, we tiled the highly homologous *HBG* proximal promoters using adenine and cytosine base editors that avoid the generation of large deletions and identified novel regulatory regions including a cluster at the –123 region. Base editing at –123 and –124 bp of *HBG* promoter induced fetal hemoglobin (HbF) to a higher level than disruption of well-known BCL11A binding site in erythroblasts derived from human CD34+ hematopoietic stem and progenitor cells (HSPC). We further demonstrated in vitro that the introduction of –123T > C and –124T > C HPFH-like mutations drives gamma-globin expression by creating a de novo binding site for KLF1. Overall, our findings shed light on so far unknown regulatory elements within the *HBG* promoter and identified additional targets for therapeutic upregulation of fetal hemoglobin.

Editor's evaluation

This paper describes the innovative use of base editing to mutagenize an enhancer region in the iconic globin locus, demonstrating a new method while also finding a potential novel locus for downstream therapeutic approaches.

Introduction

Fetal hemoglobin (HbF) is a tetramer consisting of two alpha-globin chains and two gamma-globin chains, which are highly expressed during the fetal stage of human life. The expression of HbF is silenced progressively after birth until it constitutes only about 1% of total hemoglobin (***Bauer and Orkin, 2012***). Naturally occurring mutations in the regulatory regions of the gamma-globin (*HBG*) genes have been shown to reactivate expression and increase HbF levels during adult life (***Jacob and Raper, 1958***). This inherited genetic condition is benign and is known as hereditary persistence of fetal hemoglobin (HPFH). Individuals, who inherit HPFH alongside other genetic disorders affecting the adult beta-globin gene, such as sickle cell disease or beta-thalassemia, were shown to have fewer, if any, symptoms (***Jacob and Raper, 1958***; ***Thein, 2018***). Hence, high levels of HbF expression have been shown to be beneficial for improving the clinical outcomes of patients with sickle cell anemia and beta-thalassemia.

Genome editing approaches have largely focused on the beneficial effects of HPFH mutations to increase HbF levels in sickle cell disease (***Traxler et al., 2016***). These mutations either create de novo binding sites for erythroid activators or disrupt the binding sites of repressors, thereby increasing the expression of HbF. For example, the –175T > C, –198T > C and –113A > G HPFH point mutations create de novo binding sites for the erythroid master regulators TAL1, KLF1, and GATA1, respectively (***Fischer and Nowock, 1990***; ***Martyn et al., 2019***; ***Stoming et al., 1989***; ***Wienert et al., 2017***; ***Wienert et al., 2015***). Similarly, the introduction of HPFH-associated mutations around –115 bp from the transcription start site (TSS) of *HBG* (–114C > A, –117G > A, and a 13 bp deletion [Δ13 bp]), and around –200 bp from the TSS (–195C > G, –196C > T, –197C > T, –201C > T and –202C > T/G), were shown to disrupt the binding sites of the two major fetal globin repressors, BCL11A and ZBTB7A/LRF, respectively (***Martyn et al., 2018***). However, the roles and locations of other regulatory elements in the *HBG* promoter that are involved in activation or de-repression are less well understood. Thus, tiling the *HBG* promoter using base editors could unravel molecular mechanisms of human hemoglobin switching and reveal additional point mutations that could be useful for therapeutic gamma-globin upregulation.

Targeted introduction of HPFH mutations into the *HBG* promoter by nuclease-mediated homology-directed repair is relatively inefficient and can result in high rates of random insertions and deletions (indels) through non-homologous end-joining DNA repair pathways (***Cavazzana et al., 2017***). In addition, due to the high homology between the duplicated *HBG1* and *HBG2* genes, simultaneous editing of both *HBG* promoters by programmable nucleases that cause double-stranded breaks (DSBs) sometimes results in ~4.9 kb deletion comprising the *HBG* intergenic region with uncertain consequences (***Li et al., 2018***; ***Métais et al., 2019***; ***Wienert et al., 2017***).

To overcome these limitations, we implemented a strategy to screen and identify potential regulatory mutations within the proximal promoters of the two human fetal globin genes *HBG1* and *HBG2*. We employed CRISPR base editing to introduce an array of point mutations into the *HBG* promoters and then screen for those mutations that induce HbF to therapeutic levels, without the confounding effects of creating DSBs. Similar to previous findings, we observed that base editing using adenine and cytosine base editors (ABEs and CBEs, respectively) is highly efficient in creating point mutations without inducing high levels of indels (***Gaudelli et al., 2017***; ***Komor et al., 2016***). We identified several novel point mutations that are associated with a significant increase in gamma-globin expression and could be of therapeutic interest. Our results demonstrated that base editors are a powerful tool for mapping the so far unknown regulatory elements within the *HBG* promoters and provide a proof-of-concept approach for the treatment of beta-hemoglobinopathies.

Results

Previous studies have shown that the highly homologous *HBG1* and *HBG2* proximal promoters play a crucial role in the gamma-globin expression. Several non-deletional forms of HPFH-associated point

mutations in the promoter region of *HBG1* and *HBG2* have been associated with increased expression of gamma-globin (***Wienert et al., 2015***). To identify novel regulatory elements in the human *HBG* promoters that influence gamma-globin expression, we performed a base editing screen to introduce point mutations in all compatible locations within 320 bp upstream of the TSS of the *HBG* genes. In brief, we created stable HUDEP-2 cells (***Kurita et al., 2013***) (an immortalized human erythroid progenitor cell line) expressing base editors (ABE or CBE), and then screened guide RNAs (gRNAs) targeting the proximal promoter region of *HBG1* and *HBG2* for their ability to upregulate fetal globin expression. The top gRNAs were validated for editing efficiency and HbF levels. Moreover, the plausible mechanism of novel gRNAs identified from the study on HbF elevation was further characterized by electrophoretic mobility shift assay (EMSA) and Chromatin immunoprecipitation quantitative PCR (ChIP-qPCR). Finally, the potential therapeutic induction of HbF levels for the identified novel gRNAs were validated in erythroid cells derived from healthy donor CD34+ HSPCs.

Base editors as a preferred genome editing tool for targeting the highly homologous *HBG* promoter region

First, we generated stable HUDEP-2 cell lines that express different gene editors, ABE, CBE, or Cas9, respectively. HUDEP-2 cells were transduced with ABE7.10 RA, BE3RA-FNLS, or Cas9 lentiviral constructs (hereafter named HUDEP-2-ABE, CBE, or Cas9). The vector copy number (VCN) of HUDEP-2 cells transduced with ABE, CBE, or Cas9 lentiviral constructs ranged from 0.25 to 0.85 by real-time PCR (***Figure 1—figure supplement 1a***). Previously defined sgRNAs targeting the BCL11A binding motif (***Traxler et al., 2016***) in the *HBG1* and *HBG2* promoters with a suitable editing window for ABE, CBE, and Cas9 were transduced with the VCN of 0.6–1.2 (***Figure 1—figure supplement 1a***). The editing efficiency was 88% for Cas9, 10–51% for ABE, and 59–73% for CBE, with the transduction efficiency (as measured by GFP expression) greater than 98% (***Figure 1a***). After differentiating the cells into erythroid progenitors, the percentage of HbF positive cells was higher in case of ABE and CBE than Cas9 (***Figure 1b***). While gene editing was very high with Cas9, we did not observe a corresponding increase in the HbF positive cells which might be due to a previously described 4.9 kb deletion comprising the *HBG2* gene and *HBG1-HBG2* intergenic region. It has been previously reported that introducing DSBs in highly homologous *HBG* promoters with Cas9 nucleases may generate this deletion (***Li et al., 2018***). Therefore, we determined the frequency of this deletions by qRT-PCR in all the edited samples. As expected, the 4.9 kb deletion was observed at a high frequency (76%) in Cas9 edited cells. We also noted some deletions in base edited samples but significantly fewer than with Cas9 (***Figure 1c***). Consistent with the frequency of the 4.9 kb deletion, the globin chain analysis by RP-HPLC showed lower levels of G gamma but not A gamma chain only in Cas9 edited cells in comparison to ABE and CBE edited cells after normalizing to the control (***Figure 1d***). These results suggest that ABE and CBE are highly efficient in editing the highly homologous regions like gamma-globin promoter without causing a large deletion between the two *HBG* genes.

Screening of *HBG* proximal promoter with base editors identifies novel HPFH like mutations

To identify the potential regulatory regions involved in gamma-globin expression, we therefore selected ABE and CBE for tiling the *HBG* promoter. Transcriptomic analysis of the stable cell lines expressing ABE and CBE showed a significant correlation with the wild type HUDEP-2 cells, confirming that the gene expression profiles are not altered (***Figure 1—figure supplement 1b***). As the base editors and gRNAs are constitutively expressed, we determined the editing frequency of ABE and CBE stables with gRNA-2 for its effect on HbF elevation at different time points during expansion and differentiation. The editing efficiency and HbF levels in both ABE and CBE increases over time with no discernable effect on erythroid differentiation (***Figure 1—figure supplement 1c-h***). We then generated ABE and CBE gRNAs in all compatible locations up to 320 bp upstream of the TSS of the *HBG1* and *HBG2* promoters. Guide RNAs were designed with a suitable base editing window (target nucleotide in positions 3–9 from NGG PAM distal end) for ABE and CBE (***Figure 2a***). Among the 41 gRNAs designed, 36 gRNAs had a base editing window for ABE, and 32 gRNAs had a base editing window for CBE (***Supplementary file 1***). An overview of the methodology used in this study is illustrated in ***Figure 2b***: All the gRNAs were cloned in a lentiviral vector with a GFP reporter. Lentivirus was produced for each gRNA; HUDEP-2-ABE and -CBE cells were then transduced in an arrayed format

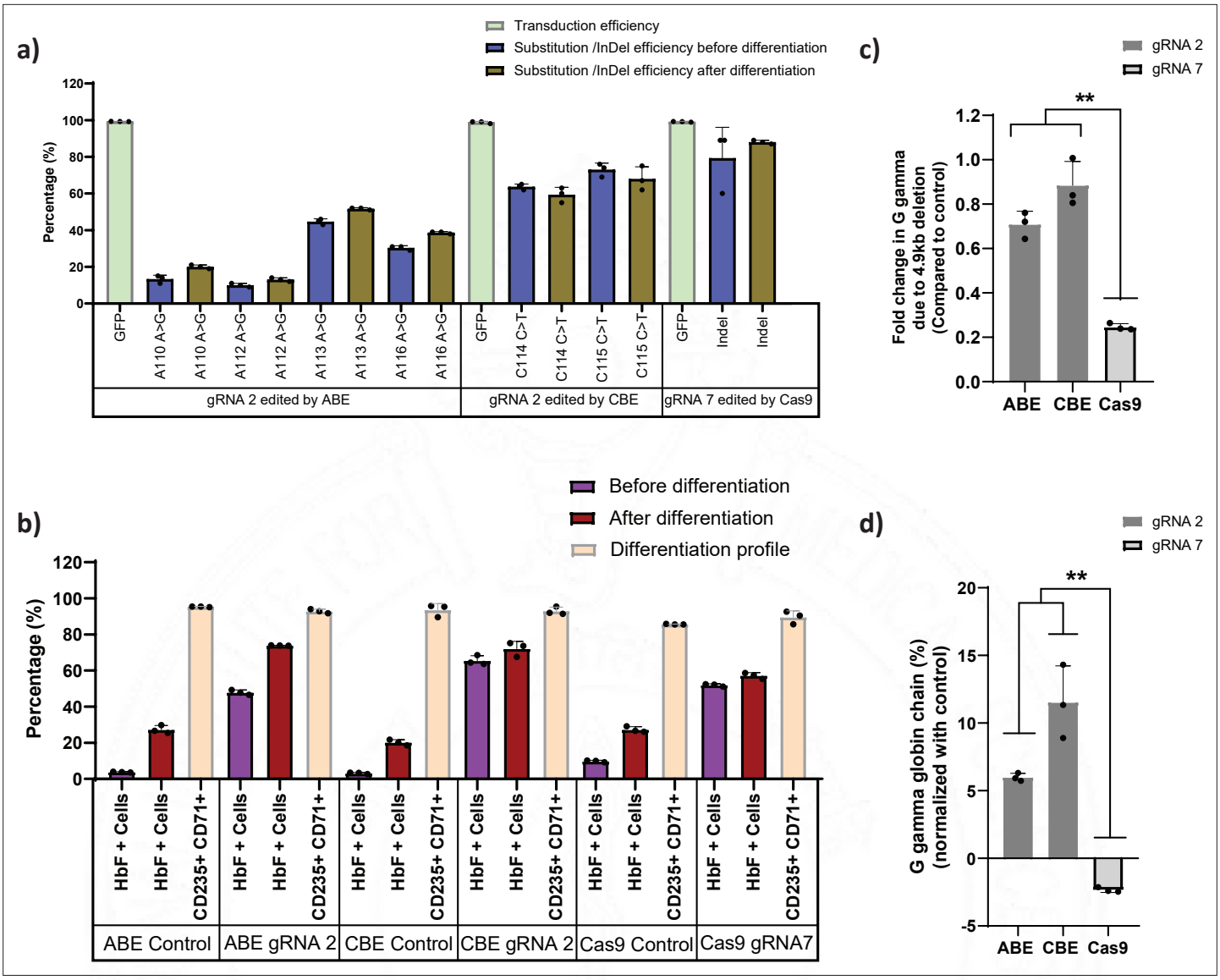


Figure 1. Base editors are preferred tool over Cas9 for editing the highly homologous *HBG1* and *HBG2* promoter. Highly homologous *HBG* promoter was edited by adenine base editor (ABE), cytosine base editor (CBE), and Cas9 with suitable guide RNAs (gRNAs) that target the well-known BCL11A binding site (–115 transcription start site [TSS]). (**a**) Transduction efficiency of gRNA-2 (for ABE and CBE) or gRNA-7 (for Cas9), percentage of individual base conversion for ABE and CBE (with gRNA-2) and insertions and deletions (indels) for Cas9 (with gRNA-7) before and after erythroid differentiation are represented. The transduction efficiency was analyzed by FACS, the individual base substitution and indel percentage were analyzed by EditR and ICE software respectively after sanger sequencing. (**b**) Flow cytometry analysis of fetal hemoglobin (HbF) and erythroid maturation markers (CD235a and CD71) expression in edited HUDEP-2 cells. The percentage of HbF-expressing cells were analyzed before and after differentiation into erythroblasts. (**c**) Analysis of *HBG2* deletion (due to 4.9 kb deletion) by qRT-PCR in the base edited and Cas9 edited HUDEP-2 cells. (**d**) Expression of G gamma-globin chain in ABE, CBE, and Cas9 edited HUDEP-2 cells, measured by RP-HPLC after differentiation into erythroblasts. The data were normalized with respective controls. Data are expressed as mean ± SEM from three biological replicates, asterisks indicate levels of statistical significance (**p < 0.01).

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Characterization of the HUDEP-2 cells expressing adenine (ABE) and cytosine base editor (CBE).

with equal transduction efficiency (~1 VCN/cell). The mean transduction efficiency for all these gRNAs in both ABE and CBE samples were around 97%. The gRNA transduced cells were then expanded for 8 days, successful base editing was then confirmed by NGS and Sanger sequencing.

First, we determined the overall efficiency of ABE-induced A-to-G conversions and CBE induced C-to-T conversions at different target sites of *HBG* promoter by NGS. The gRNAs associated with lower base editing efficiency (<10%) were excluded from further analysis as they do not provide

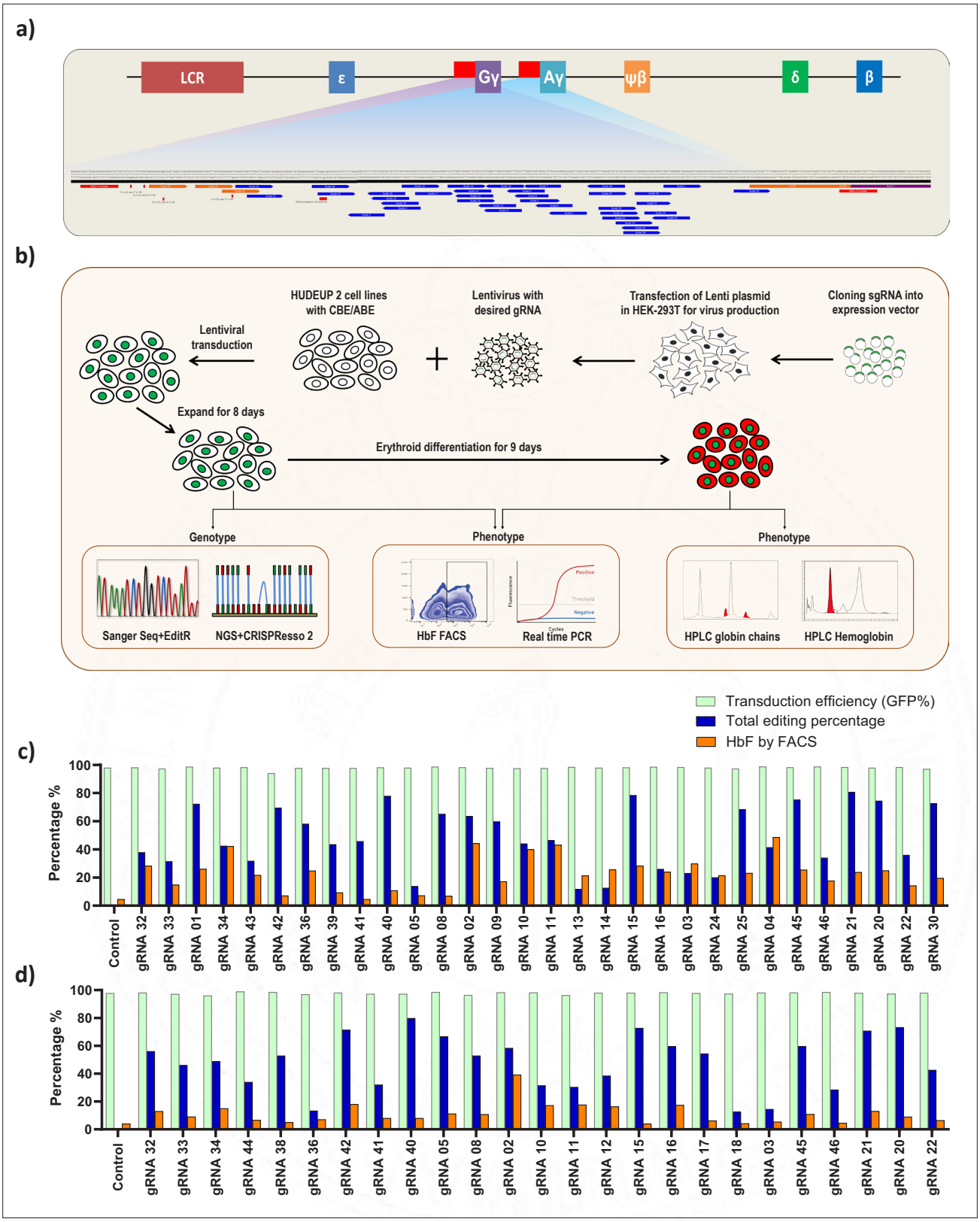


Figure 2. Screening of *HBG* promoter using base editors to identify novel point mutations that elevate fetal hemoglobin (HbF) expression. (**a**) Schematic representation of the overall screening approach, adenine base editor (ABE) or cytosine base editor (CBE) expressing HUDEP-2 cells were transduced with guide RNA (gRNAs) that target the proximal promoter of the *HBG* gene. The edited cells were expanded for 8 days. Editing efficiency was evaluated by Sanger sequencing and NGS, while functional analysis was carried out using FACS and qRT-PCR. Top targets from both the ABE

Figure 2 continued on next page

Figure 2 continued

and CBE screens with the highest induction of HbF were validated and differentiated to erythroid cells. The differentiated cells were further subjected to FACS, qRT-PCR, RP-HPLC, and HPLC analysis to determine the number of HbF positive cells, *HBG* expression, individual gamma-globin chains, and fetal hemoglobin levels, respectively. (**b**) Representation of gRNA targeting *HBG* promoter region in HUDEP-2 cell line, gRNAs targeting –320 bp upstream of transcription start site (TSS) in *HBG* genes (*HBG1* and *HBG2*) promoter regions are represented in the figure. gRNAs common for *HBG1* and *HBG2* promoters are represented in blue, while the gRNAs specific to *HBG1* promoter are represented in orange color, the primers used for deep sequencing are represented as a red bar. Comparison of transduction efficiency, base editing frequency, and HbF expression in HUDEP-2 cells expressing ABE (**c**) and CBE (**d**) transduced with different gRNAs targeting the *HBG* proximal promoter. The base edited cells were sequenced by NGS and analyzed for total editing frequency using CRISPResso-2. The transduction efficiency (GFP+ cells) and HbF positive cells were analyzed by FACS.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Analysis of base substitution efficiency at single base pair resolution in *HBG* promoter by adenine base editor (ABE) and cytosine base editor (CBE) through NGS and Sanger sequencing.

Figure supplement 2. The product purity and preferred editing window of adenine (ABE) and cytosine base editors (CBE) at the target site.

Figure supplement 3. Base editing of *HBG* promoter to identify nucleotide substitutions that suppress fetal hemoglobin (HbF) expression.

insights on *HBG* regulation (gRNAs -37, -38, -7, -18, -19, -29 in ABE, and gRNAs -1, -35, -37, -6, -7, -13, -19 in CBE) (data not shown). After excluding the low editing gRNAs, the total base editing efficiency (overall conversion achieved by a gRNA) varied from 12% to 81% and 13% to 80% for ABE (n = 30) and CBE (n = 25) respectively (***Figure 2c and d***) as determined by CRISPResso-2 analysis. The individual base conversion frequency (base conversion at single base pair resolution) of ABE (A:T to G:C) and CBE (C:G to T:A) ranged from 0% to 74% and 0% to 61%, respectively (***Figure 2—figure supplement 1a and c***). Sanger sequencing data analyzed by EditR further confirmed the base substitution efficiency at the target loci (***Figure 2—figure supplement 1b and d***). The base substitution efficiency for each gRNAs varied drastically depending on the base editing window for ABE and CBE. The average editing efficiency observed was high (>30%) in the canonical positions for ABE (A5-A7) and CBE (C5-C7), while it varied between 1% to 27% in the non-canonical positions for ABE (A1-A4, A8-A12) and CBE (C1-C4, C8-C16) for the different gRNAs used in this study (***Figure 2—figure supplement 2e-f***).

Subsequently, we explored the product purity and indel percentage for all gRNAs in ABE and CBE, as previous studies have shown that the base editors generate a low frequency of unintended edits at the target sites (***Koblan et al., 2018***). Most of the gRNAs in CBE transduced cells showed unanticipated C- to non-T edits (C-R/G-Y), among which C-to-G conversion was predominant. In the case of ABE, we observed a minimal level of unexpected base conversions (A-Y/T-R) at a few on-target sites, consistent with previous studies (***Gaudelli et al., 2017***; ***Komor et al., 2016***, ***Figure 2—figure supplement 2a-b***). The indel frequency obtained from deep sequencing data was less than 2% in both ABE and CBE (***Figure 2—figure supplement 2c-d***). Our results suggest that ABE exhibits higher product purity and lower indel frequency than CBE in all cases. In summary, we show that both ABE and CBE can effectively introduce A-to-G and C-to-T nucleotide substitutions respectively in the proximal promoters of *HBG1* and *HBG2*.

To evaluate whether the targeted base substitution at the *HBG* promoter by using ABE or CBE has increased HbF expression, we analyzed the HbF positive cells in ABE and CBE edited cells by flow cytometry after intracellular HbF staining. The percentage of HbF positive cells ranged from 2% to 44% in ABE and 1% to 35% in CBE (***Figure 2c–d***). In our preliminary analysis, among the 30 gRNAs in ABE and 25 gRNAs in CBE that we have screened for editing the *HBG* promoter region, five gRNAs in ABE and one gRNA in CBE showed a greater increase in the number of HbF positive cells (in a range of 40–50%).

We identified several gRNAs in ABE (gRNA -39, -41, -42, -08) and in CBE (gRNA -33, -44, -38, -41, -40, -15, -17, -46, -20, -21) which have higher total editing efficiency (>40%) at the target site but resulted in low HbF level (<10%). We were curious to know whether these gRNAs can affect the binding sites of activators resulting in downregulation of gamma-globin expression. To test this hypothesis, we transduced the selected gRNAs in K562 cell lines stably expressing ABE or CBE, a cell model which has high basal level of HbF expression. The transduction efficiency was more than 98% and achieved a higher individual base editing efficiency for each of the gRNAs in both ABE and CBE (***Figure 2—figure supplement 3a-b***). However, we did not observe any decrease in the number of HbF positive cells (98% of cells were HbF positive) in any of the samples suggesting that the targeted

regions did not have binding sites for essential transcriptional activators (***Figure 2—figure supplement 3c-d***).

Interestingly, some of the top candidates from the screen include target regions that were previously identified as binding sites for BCL11A (gRNA-2), KLF-1 (gRNA-4), and TAL-1 (gRNA-3), but we also identified a few other novel target sites (gRNA-10, gRNA-11, gRNA-15, gRNA-16, gRNA-21, gRNA-32, gRNA-34, gRNA-42). The gRNAs-2, -3, and -4 recreates the well-known naturally occurring HPFH mutations –114C > T, –117G > A, –175T > C and –198T > C (***Liu et al., 2018***; ***Martyn et al., 2018***; ***Stoming et al., 1989***; ***Wienert et al., 2018***; ***Wienert et al., 2017***). We compared the percentage of HbF positive cells with the editing efficiency for each of gRNAs at the target region in both ABE and CBE cells (***Figure 2c–d***). The total base editing efficiency was generally higher when compared to the proportion of HbF positive cells except in few cases (gRNA-3, -4, -13, -14, -37, and -38 in ABE edited cells). Together, the candidate gRNAs which upregulated HbF from the primary screening of the *HBG* promoter by ABE and CBE provides targets for further validation.

Base editing at potential target sites in the *HBG* promoter substantially induces HbF expression

The top eight gRNAs from the ABE screen (gRNAs -2, -3, -4, -10, -11, -15, -32, and -34) and the CBE screen (gRNAs -2, -10, -11, -16, -21, -32, -34, and -42) which resulted in the highest levels of HbF positive cells were further validated. Out of the top eight gRNAs identified from the base editor screen, five gRNAs (gRNA-2, gRNA-10, gRNA-11, gRNA-32, and gRNA-34) were common in both ABE and CBE, indicating that these target regions might play an important role in *HBG* silencing. The edited cells were cultured in erythroid differentiation media after the initial expansion, and a set of functional assays were carried out (***Figure 2b***). Corresponding to the screening results, the total editing efficiency ranged from 24% to 78% and 36% to 85% with mean transduction efficiencies of 96% and 90% for ABE and CBE, respectively (***Figure 3a–b*** and ***Figure 3—figure supplement 1f***). We observed individual base conversion of A- to-G (ranging from 0% to 65%) or C- to -T (ranging from 1% to 57%) at the respective target regions with less than 2% indel frequency (***Figure 3—figure supplement 1a-b and e***). Further, we also observed the undesired non-C-to-T conversions (i.e., C- to-A or C-to-G) at the on-target site by CBE but not with ABE (***Figure 3—figure supplement 1a-b*** and ***Figure 3—figure supplement 2a-b***). The distribution of specific nucleotide substitution mediated by ABE or CBE for all the top eight gRNAs are highlighted in ***Figure 3—figure supplement 2a-b***, respectively. ABE showed higher base editing efficiencies of the cognate A and Ts (A113 and A116 for gRNA-02, T175 for gRNA-03, T198 for gRNA-04) than the bystander A and Ts (A110 and A112 for gRNA-02, T181 for gRNA-03, T199 for gRNA-04) for the creation of HPFH mutations. In the case of CBE, we also observed the C-to-T base conversion at the nucleotides adjacent to the protospacer sequence as previously observed (***Arbab et al., 2020***; ***Webber et al., 2019***). One such example is gRNA-10 and gRNA-11 in CBE; we observed the base conversion outside the protospacer sequence (–117 site) in addition to on-target editing at –122 site within the base editing window (***Figure 3—figure supplement 1b*** and ***Figure 3—figure supplement 2b***). The base conversion at –117 site disrupts the core binding motif of the major fetal globin repressor – BCL11A (***Martyn et al., 2018***; ***Wienert et al., 2018***; ***Yang et al., 2019***). We distinguished the editing frequency in *HBG1* and *HBG2* promoters by phasing the edits with single nucleotide variations at positions –271, –307, –317, and –324 which are unique in *HBG1* and *HBG2* promoters, using Bowtie 2 and IGV software (***Robinson, 2012***; ***Langmead and Salzberg, 2013***). Our analysis showed that base editing rates were highly similar and there is no variation in base substitution efficiency between the highly homologous *HBG1* and *HBG2* promoters (***Figure 3—figure supplement 1c-d***).

We analyzed the *HBG* expression before and during differentiation by qRT-PCR. We observed a significant increase in the *HBG* mRNA expression for all the top eight gRNAs in ABE edited cells ($p < 0.01$ - $p < 0.0001$) (***Figure 3—figure supplement 3a***). In the case of CBE, gRNAs -2, -10, -11, and-42 showed a substantial increase in *HBG* mRNA expression ($p < 0.05$ - $p < 0.0001$), while gRNAs -16, -21, -32, and -34 showed a modest level of expression as compared with the control (***Figure 3—figure supplement 3g***) before differentiation. The globin mRNA expression pattern in both ABE (***Figure 3—figure supplement 3b***) and CBE (***Figure 3—figure supplement 3h***) edited cells also followed a similar trend during erythroid differentiation. We also determined the number of HbF positive cells before and after erythroid differentiation using FACS. As expected, the percentage of HbF positive

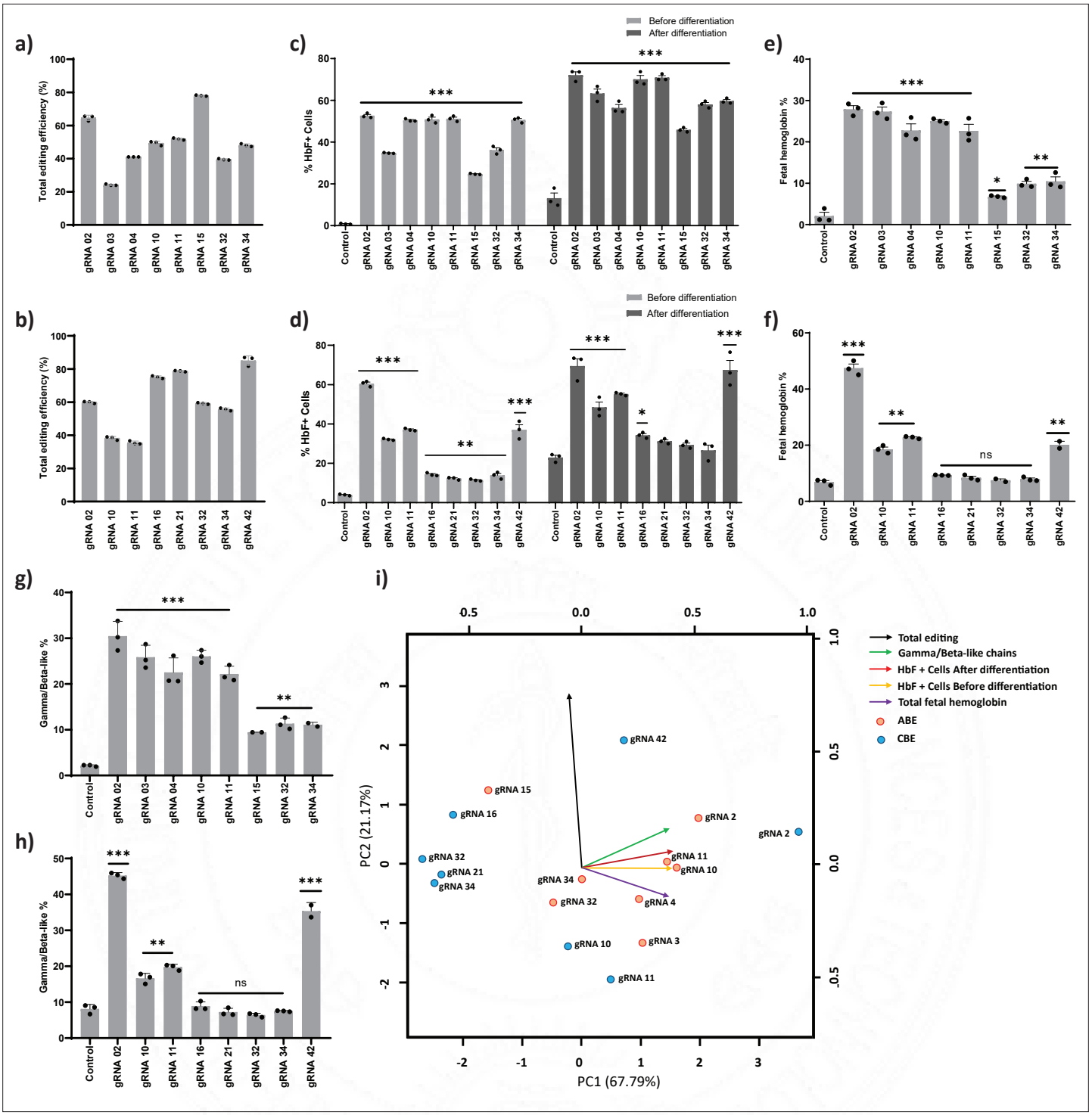


Figure 3. Validation of targeted base editing for top eight guide RNAs (gRNAs) from the primary screen of adenine base editor (ABE) and cytosine base editor (CBE) at *HBG* promoters. HUDEP-2 cells expressing ABE or CBE transduced with the top eight gRNAs were analyzed by deep sequencing at the targeted regions in the *HBG* promoter. The total editing efficiencies of ABE (**a**) or CBE (**b**) are represented as the percentage of total sequencing reads with target C:A converted to T:G at specified sites. Evaluation of fetal hemoglobin (HbF) positive cells in HUDEP-2 cells expressing ABE (**c**) or CBE (**d**) transduced with the respective gRNAs, before and after differentiation by flow cytometry; globin chains analysis in ABE (**e**) or CBE (**f**) edited HUDEP-2 cells after erythroid differentiation by RP-HPLC; HbF analysis in ABE (**g**) or CBE (**h**) edited HUDEP-2 cells after erythroid differentiation by HPLC. (**i**) Principal component analysis plot for the correlation between the outcomes of base editing using top eight gRNAs. The relationship between the base edit frequency, HbF+ cells, HbF, and gamma/beta-like chains in ABE or CBE edited HUDEP-2 stable cell for the indicated gRNAs were analyzed.

Figure 3 continued on next page

Figure 3 continued

The first two principal components are plotted, and the variance accounted for by each principal component is shown. Data are expressed as mean ± SEM from three biological replicates ($p > 0.05$). Asterisks indicate levels of statistical significance **$p < 0.01$, ***$p < 0.001$.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Assessment of base editing efficiency at the highly homologous *HBG1* and *HBG2* promoter.

Figure supplement 2. Summary of alleles frequency for top eight guide RNAs (gRNAs) at target site by adenine base editor (ABE) and cytosine base editor (CBE).

Figure supplement 3. Adenine and cytosine base editing of *HBG* promoter on globin chain mRNA and protein expression.

Figure supplement 4. Evaluation of the low efficiency guide RNAs (gRNAs) that induced fetal hemoglobin (HbF) with hyperactive variant adenine base editor (ABE)8e .

cells in differentiated erythroid cells was slightly higher than that of the undifferentiated edited cells (***Figure 3c–d***). Further, we determined the effect of base editing on erythroid differentiation using flow cytometry analysis with CD235a and CD71 markers. The shift in expression of CD71 positive cells alone to CD71/CD235a double positive cells reflects the erythroid differentiation pattern of HUDEP-2 cells (***Kurita et al., 2013***). The percentage of double positive cells was 83–90% for CBE and above 95% for ABE edited cells compared to the control, which was 77% and 97%, respectively, suggesting that the differentiation ability of the edited cells was not affected (***Figure 3—figure supplement 3d*** and ***Figure 3—figure supplement 3j***).

Furthermore, the level of globin chains was analyzed by using reverse-phase HPLC in differentiated erythroid cells from both ABE (***Figure 3—figure supplement 3c***) and CBE edited samples (***Figure 3—figure supplement 3i***). We observed a significant induction of gamma-globin chain expression, which represented 10% to30% of total beta-like globin content in all the ABE edited samples (gRNAs -2, -3, -4, -10, -11, -15, -32, and -34) (***Figure 3g***). In the case of CBE, the gamma-globin chain levels were around 6% to45% of total beta-like globin content, among which gRNAs -2, -42, -10, and -11 showed significant elevation when compared to the control (***Figure 3h***). In both ABE and CBE edited cells, the increase in gamma-globin chains was consistently associated with a reciprocal reduction in beta-globin chains thereby maintaining the alpha to beta-like globin chain ratio (***Figure 3—figure supplement 3c*** and ***Figure 3—figure supplement 3i***). Even though the *HBG1* and *HBG2* promoters were base edited with equal efficiency, *HBG1* showed moderately higher expression levels compared to *HBG2* in most of the ABE and CBE edited cells. The decrease in *HBG2* expression in these samples might be due to biased HbF regulation or the 4.9 kb deletion that deletes the *HBG2* gene.

To find whether decrease in *HBG2* expression is due to the 4.9 kb deletion, gRNAs which showed significant reduction in *HBG2* over *HBG1* expression in ABE (gRNA-4, -10, -11, -15, -32, -34) and CBE (gRNA-2, -10, -11, -32, -34, -42) edited cells were further investigated by qRT-PCR. Interestingly, the frequency of 4.9 kb deletion in ABE ranged from 2% to 32% while in CBE it ranged from 0% to 12% (***Figure 3—figure supplement 3e*** and ***Figure 3—figure supplement 3k***). To determine the correlation between the reduction in *HBG2* chain expression and the frequency of large deletion, Pearson correlation analysis was performed in the above-mentioned gRNAs in ABE and CBE edited cells. We observed a high correlation ($r = 0.71$) in the case of ABE, whereas much lower correlation was observed with CBE ($r = 0.26$) (***Figure 3—figure supplement 3f*** and ***Figure 3—figure supplement 3i***). These data suggest that the reduction in G gamma chain expression is due to higher frequency of deletions in the ABE edited samples, while in the case of CBE, the decrease in the G gamma chain expression is independent of larger deletions and might be due to the biased expression of gamma-globin.

We observed a substantial difference in deletion rates across gRNAs as well as between base editors (***Figure 3—figure supplement 3e*** and ***Figure 3—figure supplement 3k***). The difference in the DNA sequence composition which often affects the editing efficiency might also be responsible for varied deletion observed across the gRNAs targeting the *HBG* promoter. On the other hand, processivity of the editors could account for the difference in deletion observed with the same gRNA while editing with CBE and ABE. As the base editors cannot dock on an already edited strand, CBE with a higher rate of editing (~50% editing on day 1) is prevented from interacting again with the DNA, thus reducing the chances of deletion compared to ABE which takes longer to achieve similar editing (~50% editing on day 8) (***Figure 1—figure supplement 1b and c***). This observation is further supported by the minimal deletion seen in samples edited with ABE8e which has a higher processivity

(~90% editing within 24 hr) compared to both CBE and ABE 7.10 (***Figure 3—figure supplement 4c***, ***Richter et al., 2020***).

Next, we analyzed the level of hemoglobin tetramers in the differentiated cells by HPLC to determine whether increase in *HBG* chain expression resulted in functional HbF production. We observed a significant induction of HbF in all the top eight target gRNA transduced cells in ABE (gRNAs- 2, -3, -4, -10, -11, -15, -32, and -34), whereas only gRNA-2, -10, -11, and -42 expressed higher levels of HbF in CBE edited cells (***Figure 3e–f***). Consistent with globin chain analysis, we observed that the increase in HbF variant is associated with compensatory downregulation of adult hemoglobin levels in the edited cells. The relationship between the editing efficiency and HbF expression was analyzed for all the top-scoring gRNAs in ABE and CBE edited cells (***Figure 3i***). Among the validated top eight gRNAs in ABE and CBE, gRNA-2, -10, -11 with ABE and gRNA-2 with CBE resulted in a high target editing efficiency with a corresponding increase in HbF expression. In case of gRNA-42 with CBE, only a modest level of HbF elevation was achieved even with higher editing efficiency. On the other hand, gRNAs -3 and -4 with ABE and gRNAs -10 and -11 with CBE showed higher elevation of HbF levels despite lower base conversion efficiency. The higher number of HbF positive cells with minimal base editing might be due to heterogenous editing at target site per cell since there are two copies each of *HBG1* and *HBG2*. Further, irrespective of editing at the target region, the binding of CRISPR-Cas9 complex through the gRNAs at the *HBG* promoter might disrupt the binding of major transcriptional repressor that are involved in globin expression (***Shariati et al., 2019***).

Among the samples which resulted in higher HbF induction with lower editing efficiency, we validated gRNA-03 and -11 with a hyperactive variant of ABE (ABE8e) to determine whether further elevation in HbF level can be attained by increasing the editing efficiency. The HUDEP-2 cells stably expressing ABE8e were transduced with gRNA-03 and -11, with a VCN of 0.28% for the editor and 0.56% for the gRNA (***Figure 3—figure supplement 4e***). We observed a high percentage of base substitution at the target site (more than 95%) with the corresponding increase in the HbF positive cells and gamma-globin chains in both the gRNAs (***Figure 3—figure supplement 4a-c***). The erythroid differentiation capacity of the edited cells was equivalent to that of control (***Figure 3—figure supplement 4b***). The frequency of larger deletions was also significantly reduced perhaps because of the higher processive rate of ABE8e (***Figure 3—figure supplement 4d***). Thus, the ABE8e variant has improved the base editing efficiency at the target region and provided higher level of HbF induction with a reduced frequency of larger deletion.

Through this screen, we identified multiple novel individual regulatory regions and validated well-known HPFH mutations in the *HBG* proximal promoter that are important for gamma-globin regulation. Interestingly, gRNA-2 (with ABE or CBE) and gRNA-4 (with ABE) disrupt the binding site for the major gamma-globin repressors, BCL11A and LRF/ZBTB7A, and generate the binding motif for the transcriptional activators, GATA1 and KLF1, thus resulting in overall activation of *HBG* expression. Base editing of –114C > T, –115C > T and –116A > G mutations disrupts the binding of BCL11A and the base conversion at –198T > C and –199T > C affect the binding of LRF/ZBTB7A to the *HBG* promoter. Further, the installation of –113A > G and –198T > C mutations by gRNA-2 and gRNA-4, generate a binding site for GATA1 and KLF-1, respectively. Moreover, the base conversion at −175T > C by gRNA-3 (with ABE) creates a TAL1 binding site. The novel gamma-globin regulatory region identified includes the target base substitution mediated by gRNA-10, -11, -15, -21, -32, and -34 with ABE or CBE. With gRNAs -10 and -11, ABE converts nucleotide at –123T > C and –124T > C position, whereas CBE converts nucleotide at –122G > A (on-target editing site) and –117G > A (outside the editing window) positions. Target base substitutions at –123T > C and –124T > C positions result in greater induction of *HBG* expression, equivalent to that of the known HPFH mutations that disrupt the binding of BCL11A. Overall, the potential gRNAs include gRNA-2, 3-, -4, -10, and -11 with ABE and gRNA-2 with CBE that exhibit high induction of HbF expression.

Base editing of the –123 region of *HBG* promoter in human CD34+ HSPCs

To further determine the therapeutic potential of novel targets identified from this study on induction of gamma-globin expression, we performed base editing of CD34+ HSPCs from healthy donor (***Figure 4a***). Electroporation of the ABE8e mRNA with gRNA targeting the BCL11A binding site (gRNA-2) and –123 novel cluster (gRNA-11) effectively generated highly efficient base editing at the

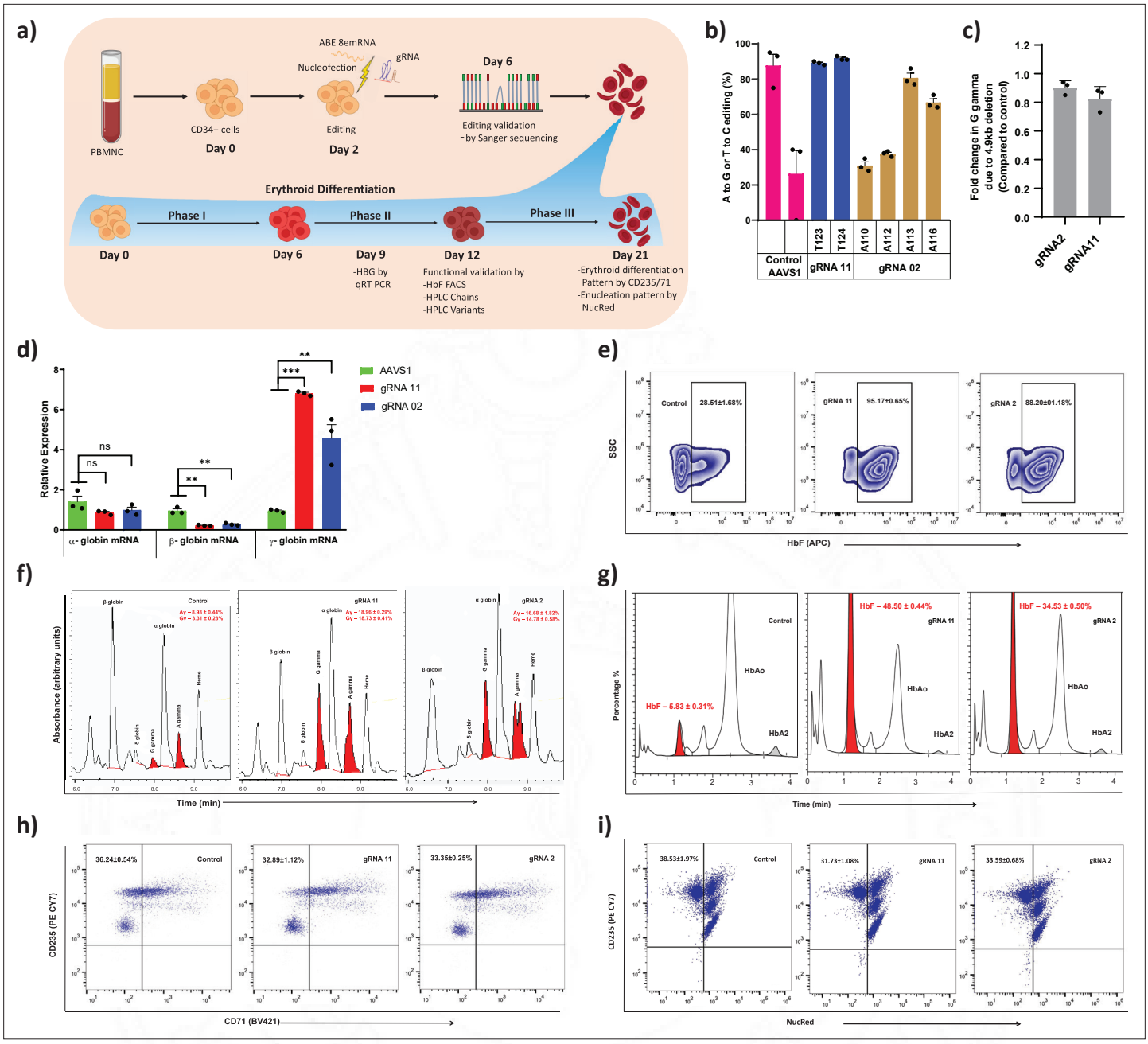


Figure 4. Therapeutic induction of fetal hemoglobin (HbF) in erythroblast derived from healthy donor CD34+ hematopoietic stem and progenitor cells (HSPCs) upon base editing of *HBG* promoter. (**a**) Schematic representation of steps involved in based editing of CD34+ HSPCs. Mobilized CD34+ HSPCs from healthy donor were nucleofected using MaxCyte system with adenine base editor (ABE)8e mRNA and respective guide RNAs (gRNAs) on day 2 of expansion. During expansion, CD34+ HSPCs were analyzed at day 6 for the editing efficiency and 4.9 kb deletion. (**b**) Efficiency of individual base conversion at the target sites were measured by EditR after Sanger sequencing. (**c**) Analysis of *HBG2* deletion (due to 4.9 kb large deletion) by qRT-PCR. The based edited CD34+ HSPCs were cultured in a three-phase liquid culture system for erythroid differentiation and enucleation. (**d**) Relative expression of globin transcripts analyzed by qRT-PCR (ΔΔCT) in erythroblasts derived from base edited CD34+ HSPCs on day 9 of differentiation. The functional validation of HbF elevation was analyzed in erythroblasts derived from the indicated samples by FACS, HPLC, and RP-HPLC on day 12 of erythroid differentiation. (**e**) HbF positive cells analyzed by flow cytometry are represented as zebra plots. (**f**) RP-HPLC chromatogram profiles of individual globin chains and (**g**) HPLC chromatogram profile of hemoglobin variants. On the final day of erythroid differentiation, the expression of maturation markers and enucleation fraction were measured by FACS analysis. (**h**) Flow cytometry for the erythroid maturation markers CD235a+ and CD71+. (**i**) Enucleation pattern was determined by flow cytometry analysis for CD235a with NucRed in erythroid cells derived from CD34+ HSPCs. Asterisks indicate levels of statistical significance $^{**}p < 0.01$, $^{***}p < 0.001$.

target site. The editing efficiency observed at individual base positions were –110 (31%), –112 (37%), –113 (80%), and –116 (66%) with gRNA-2 and –123 (89%), –124 (91%) with gRNA-11 (***Figure 4b***). In case of gRNA-11, the base editing events generated a high proportion of –123 and –124 mutations in combination at the target site. We cultured the base edited CD34+ HSPCs under erythroid differentiation conditions and analyzed HbF expression. The relative levels of *HBG* expression were significantly higher in gRNA-11 (>6-fold) and gRNA-2 (>5-fold) edited samples when compared to control (AAVS1 edited sample) by qRT-PCR (***Figure 4d***). In contrast, a significant downregulation of *HBB* and unchanged levels of *HBA* expression were observed in both the tested targets. Similarly, we observed a substantial increase in HbF protein expression in erythroblast derived from base edited CD34+ HSPCs. Flow cytometry and HPLC variant analysis confirmed the robust increase in the proportion of HbF positive cells and their HbF content compared with control samples for all the tested targets, with the higher effect in gRNA-11 (***Figure 4e and g***). The globin chain analysis showed an increase in expression of *HBG1* and *HBG2* globin chain levels and a reduction of *HBB* globin chain level (***Figure 4f***). Importantly, base editing of the *HBG* proximal promoter with gRNA-2 or -11 did not alter enucleation potential or the expression of erythroid maturation markers CD235a or CD71 (***Figure 4h–i***). Finally, we determined the frequency of the 4.9 kb deletion in CD34+ HSPCs electroporated with ABE8e and gRNA-2 or -11. We observed a very minimal frequency of the 4.9 kb deletion which might be due to higher processivity and transient expression of the base editor mRNA (***Figure 4c***). The present results suggest that the level of HbF induction mediated by the installation of novel –123 cluster HPFH-like mutations (through gRNA-11) is comparable to the naturally occurring –115 cluster HPFH mutations (through gRNA-2) that disrupt the binding site of BCL11A. Together, our data demonstrate that adenine base editing of the *HBG1* and *HBG2* promoters to recreate the novel –123 cluster HPFH-like mutations is a potential approach for the therapeutic induction of fetal globin level and treatment for beta-hemoglobinopathies.

The –123 T>C and –124 T>C HPFH-like mutations creates a de novo binding site for KLF1

Finally, we investigated the possibility that novel HPFH-like mutations introduced by the base editor might either create or disrupt the binding site for transcriptional regulators. Interestingly, we observed that the base editing at –123T > C and –124T > C sites by ABE with a single gRNA creates the consensus binding site for the master erythroid transcription factor KLF1 (***Figure 5a and b***; ***Tallack et al., 2010***). We performed EMSA to verify binding of KLF1 to a probe containing this core element. We observed modest but clear binding of KLF1 to the –123T > C and –124T > C mutated probe in EMSA but not with the wild type probe (***Figure 5—figure supplement 1a-b***) or probes containing either –123T > C or –124T > C mutations alone (***Figure 5c–d***). This confirms that the combination of –123T > C and –124T > C mutation is important for the KLF1 binding to the *HBG* promoter. Next, we performed ChIP experiments to determine whether the KLF-1 directly interacts with –123T > C and –124T > C mutated region of *HBG* promoter in vivo. The KLF1 ChIP was performed in three independent HUDEP-2 clones sorted from wild type cells or the double mutant edited cells, respectively. The ChIP results were normalized to an unrelated positive control, a KLF1 binding site at the SP1 locus. We observed a weak increase in the signal of KLF1 binding to the *HBG* promoters in the cells edited to contain the –123 and –124 mutations, but the effect was modest, and a similarly weak enhancement was also observed at an arbitrary negative control locus, VEGF (***Figure 5—figure supplement 1c***). Thus, as seen in the EMSA, the KLF1 binding is at best weak and may be below the level of detection by ChIP. Future investigations would be required to confirm that KLF1 binding to this site is the main in vivo mechanism of –123T > C and –124T > C HPFH driven upregulation of gamma-globin.

Off-target and gene expression analysis after base editing at the *HBG* promoter

ABE and CBE are known to create Cas-dependent DNA off-target and transient Cas-independent RNA off-target at low levels (***Anzalone et al., 2020***). It has been reported that the Cas-independent DNA off-target is very low and undetectable (***Anzalone et al., 2020***). We used Cas-OFFinder tool to predict the Cas-dependent DNA off-target for the novel gRNA (gRNA-11). The identified target regions were deep sequenced by NGS. Despite the higher on-target efficiency, off-target editing was not observed at the top target sites (***Figure 6a***). Next, we performed transcriptome-wide RNA

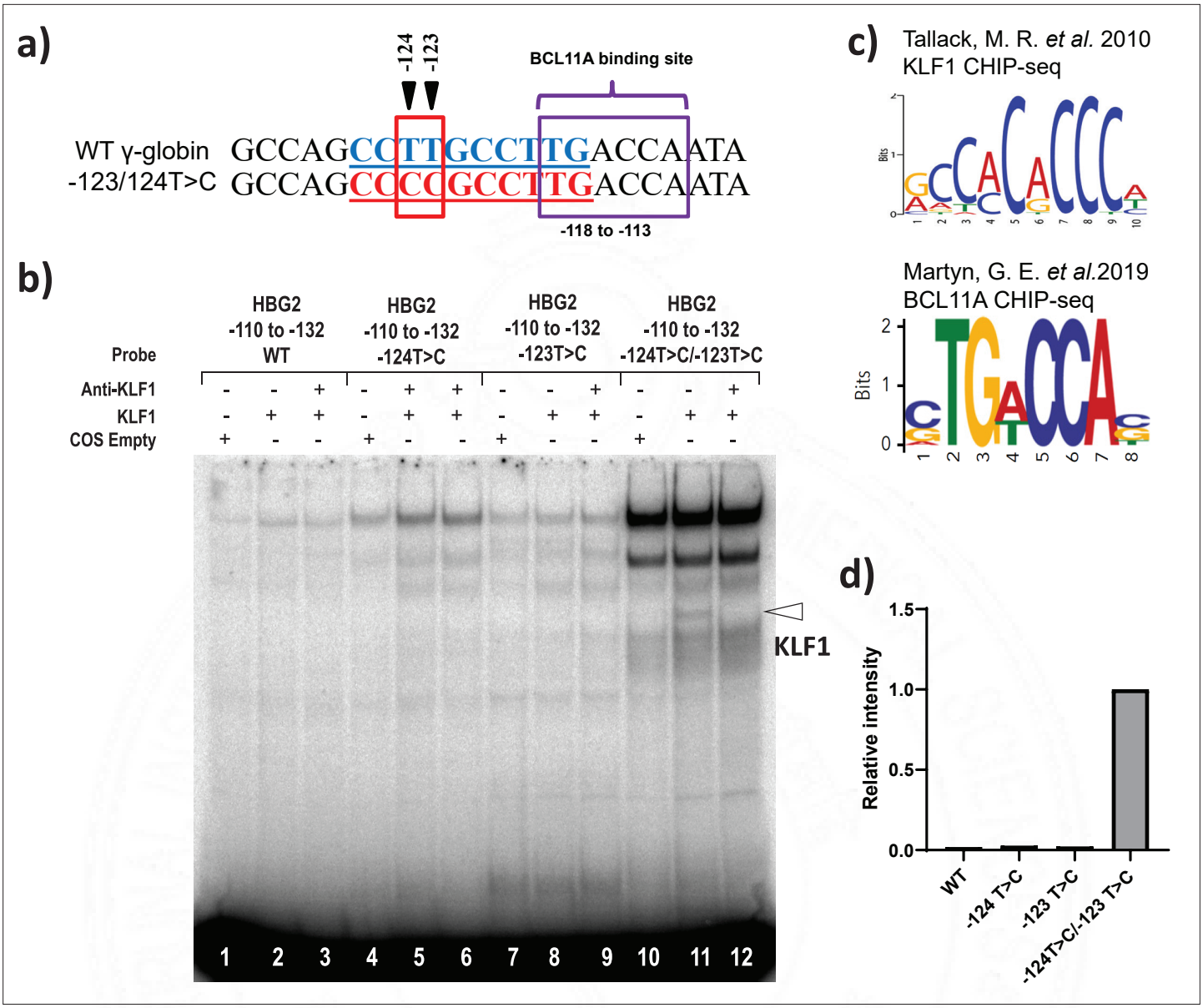


Figure 5. KLF1 binds to the –123T > C and -124T > C region of the *HBG* proximal promoter in vitro. (**a**) Introduction of T-to-C mutation at –123 and –124 of the *HBG* promoter (–132 to –110 bp) creates the de novo binding site for the KLF1, the wild type and novel KLF binding motif is highlighted in blue and red, respectively. (**b**) In vivo binding motifs of transcription factors KLF1 and BCL11A as determined by ChIP-Seq as previously reported. (**c**) Electrophoretic mobility shift assay (EMSA) showing KLF1 binding to –123T > C/–124T > C probe but failing to bind to –124T > C probe, –123T > C probe and WT probe with the –123T/–124T region of the *HBG* promoter in vitro. Lanes 1, 4,7, and 10 contain nuclear extracts from COS cells transfected with a pcDNA3 empty vector. Lanes 2–3, 5–6, 8–9, and 11–12 contain nuclear extracts from COS cells overexpressing KLF1. Binding of KLF1 to the –123T > C/–124T > C HPFH mutant probe can be observed in lane 11, with a super shift of KLF1 in the presence of anti-KLF1 antibody in lane 12. (**d**) Quantification of relative intensity of bands (KLF1 binding to the probe) from the EMSA using Image Lab 6.0.1 (Bio-Rad) software.

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. Electrophoretic mobility shift assay (EMSA) showing KLF1 binding to –123T > C/–124T > C probe but failing to bind to –124T > C probe, –123T > C probe, and wild type (WT) probe with the –123T/–124T region of the *HBG* promoter in vitro.

Figure supplement 1. Recruitment of KLF1 to site at –123 bp of *HBG* proximal promoter were analyzed by electrophoretic mobility shift assay (EMSA) and ChIP-qPCR.

Figure supplement 1—source data 1. Electrophoretic mobility shift assay (EMSA) showing the binding of KLF1 to the –123T > C/–124T > C probe but fails to bind to a wild type (WT) probe containing the –123/–124 region of the *HBG* promoter in vitro.

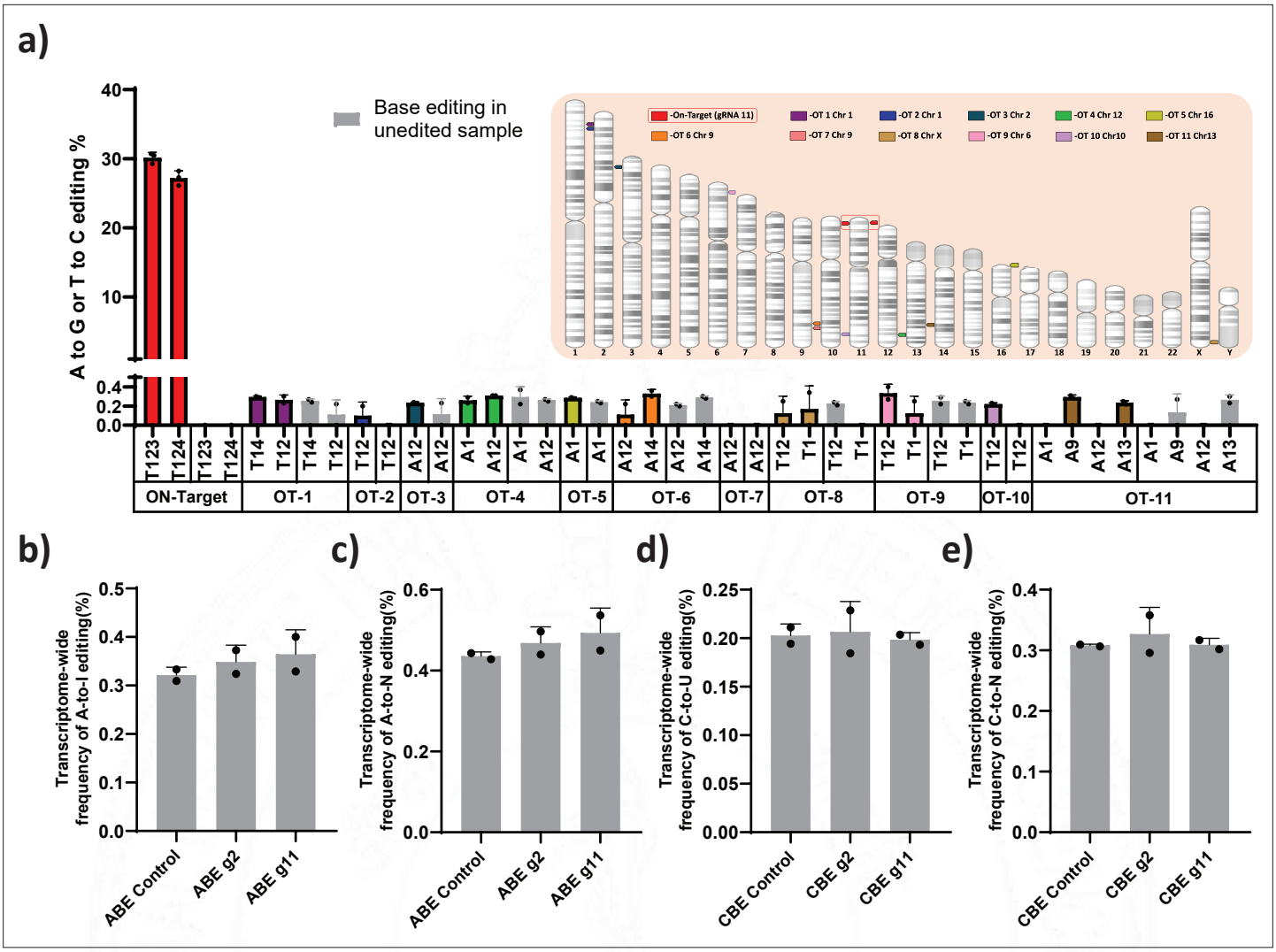


Figure 6. Evaluation of Cas-dependent DNA off-target and Cas-independent RNA off-target editing by adenine base editor. (**a**) Base conversions at the top 11 Cas-dependent DNA off-target sites in adenine base editor (ABE) 7.10 stable edited with guide RNA (gRNA)-11, along with the on-target events. The positions of the off-target and on-target loci are represented in their respective chromosome. The frequency of transcriptome-wide cellular levels of A- to- I (**b**), A- to- N (**c**), C- to -U (**d**), and C- to- N (**e**) RNA editing in BE3 stables (CBE), ABE 7.10 stables (ABE), BE3 stables edited with gRNAs-2 or -11, and ABE edited with gRNAs-2 or -11 are represented. The data are mean ± SD of two technical replicates.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Expression profile of *HBG* regulators after base editing.

sequencing on ABE and CBE stables with or without gRNA-2 and -11 to check whether base editing induced major spurious RNA deamination. The distribution frequency of A-to- I (in ABE) or C-to-U (in CBE) conversion across the base edited samples was very similar to that of the parental stable cell line (***Figure 6b–e***). To further verify that editing the gamma-globin promoter is not affecting the expression of other genes involved in globin regulation, we performed the differential analysis on these samples for the specific genes involved in globin regulation. We observed that there is no significant difference between the edited and control cells except for both gamma- and delta-globin genes (***Figure 6—figure supplement 1***).

Overall, these results support that the ABE and CBE are useful in creating specific point mutations in the homologous *HBG1* and *HBG2* promoters, leading to a potential increase in the number of HbF positive cells and overall HbF production with a significant reduction in the larger deletion frequency.

Discussion

During normal globin switching, interactions of *cis*-acting elements with several different transcription factors lead to the silencing of fetal globin and in turn the activation of beta-globin (***Ikuta et al., 1996***). To obtain insights into the regulation of gamma-globin gene expression, we have used two complementary base editing approaches to screen the *HBG* promoter at single nucleotide resolution. This approach allowed us to identify several novel nucleotide substitutions in the *HBG* promoter that elevate HbF levels by altering the binding site for transcriptional activators or repressors.

Current approaches to studying fetal globin regulation by programmable nucleases often result in the deletion of the *HBG2* gene due to the introduction of DSBs in both *HBG* promoters (***Traxler et al., 2016***). The elimination of the 4.9 kb intergenic region (including the *HBG2* gene) appears to allow the locus control region (LCR) to directly interact with the *HBG1* promoter and drive its expression (***Métais et al., 2019***). It can be challenging to determine the exact role of different HPFH mutations on individual gamma-globin expression because mutations can occur in either or both *HBG2* and *HBG1* promoters. Further, the CRISPR-Cas9-based editing produces different combination of indel at the target sites which makes it difficult to pinpoint the precise mutations involved in the gene regulation. A base editing strategy converts target bases in the editing window without the generation of DSBs and hence largely avoids splicing of the *HBG* locus. Using this strategy, we targeted regions in both *HBG1 and HBG2* promoters and were able to efficiently edit sites in the promoters with fewer or no large deletions, which gave us the opportunity to evaluate gamma-globin expression from two active promoters.

We did observe a small percentage of 4.9 kb deletions even with base editors that use a nickase variant of the CRISPR/Cas9 system in our study. The larger deletions may be mainly a result of simultaneous CRISPR-Cas9-induced DSBs or by paired nickase-mediated two single-strand breaks (SSBs) on opposite DNA strands of the *HBG1* and *HBG2* gene (***Ran et al., 2013a***). Interestingly, recent studies have shown the possibility of adjacent SSB on the same DNA strand leading to the formation of genomic deletions in plants. The deletion frequency depends upon the initial release of the single-stranded fragment between the two SSBs (***Schiml et al., 2016***). Further reports suggest the conversion of the persistent nick into DSBs by the replication fork. The R-loop primed replication fork encounters the single-strand nick site in DNA template and collapses to produce a DSB (***Kuzminov, 2001***; ***Wimberly et al., 2013***). Based on these findings, we predict the concurrent introduction of SSB by base editors at the editing site of the *HBG1* and *HBG2* promoter might generate some 4.9 kb larger deletions, though we observed very few.

Several different HPFH point mutations have been reported in the *HBG* promoters; and the effect of these mutations on gamma-globin expression in the native cellular environment has been deciphered for this limited set of mutations (***Bauer and Orkin, 2012***; ***Liu et al., 2018***; ***Martyn et al., 2019***; ***Wienert et al., 2017***; ***Wienert et al., 2015***). Our findings are in agreement with previous reports that the point mutations in three different regions of the *HBG* promoters centered around positions −198, –175, and −115 mimic the HPFH-associated point mutations affecting essential regulators of HbF expression (***Liu et al., 2018***; ***Martyn et al., 2019***; ***Martyn et al., 2018***; ***Stoming et al., 1989***; ***Wienert et al., 2017***; ***Wienert et al., 2015***). Among the known HPFH point mutations, base conversion within the –115 cluster (from –110 to –116) showed the highest increase in promoter activity, confirming previous studies (***Fucharoen et al., 1990***; ***Gilman et al., 1988***; ***Zertal-Zidani et al., 1999***; ***Motum et al., 1994***). CBE-mediated base conversion (C- to- T) at positions –114 and –115 resulted in a significantly greater induction of HbF than the multiple A-to-G nucleotide substitutions at −110, –112, −113, and –116 positions made by ABE. Recently, it has been shown that the major HbF repressor BCL11A directly binds to the core TGACC motif located at – 114 to –118 (***Liu et al., 2018***; ***Martyn et al., 2018***). Naturally occurring HPFH mutations at –117G > A, –114C > A, –114C > T, –114C > G, and Δ13bp disrupts binding of BCL11A to the promoter (***Martyn et al., 2018***). The –113A > G HPFH mutation within the –115 cluster creates a binding site for the master erythroid regulator GATA1 without disrupting the binding of BCL11A (***Martyn et al., 2019***). Our results are consistent with these previous reports showing that disruption of the core binding region of BCL11A and the creation of a de novo binding sites for GATA1 results in the elevation of fetal globin in wild type HUDEP-2 cells (***Wienert et al., 2017***). ABE-mediated T-to-C substitution at position –198 of the *HBG* gene promoter has previously been shown to be associated with British HPFH and substantially elevates expression of HbF by creating a de novo binding site for the erythroid gene activator KLF1 (***Tate et al., 1986***;

Wienert et al., 2017). Another known HPFH mutation (–175T > C) has been shown to promote enhancer looping to the *HBG* promoter through recruitment of the activator TAL1 (***Wienert et al., 2015***). Further, increased editing efficiency at the –175T > C position with the hyperactive variants of ABE (ABE8e) resulted in the highest induction of HbF synthesis in human erythroid cells.

In this study, we have identified several new point mutations in the *HBG* promoter associated with high HbF levels. *HBG* promoter base editing by ABE-mediated conversion (A- to-G) revealed multiple potential HbF regulatory regions compared to CBE since the targeted region had more ABE-compatible gRNAs than CBE. In addition to the known mutations, we have identified novel substitutions at –69 (C -to- T), –70 (C- to- T), –122 (G -to -A), –123 (T -to- C), and –124 (T -to -C) of the *HBG* promoters as potential new regulatory mutations that can elevate gamma-globin expression. The levels of gamma-globin expression resulting from these mutations were very similar to those of well-characterized, naturally occurring HPFH mutations. Our study has predicted that nucleotide substitutions at –123T > C and –124T > C positions of the *HBG* promoter might result in reactivation of gamma-globin expression through the creation of a binding site for KLF-1, which was then confirmed by EMSA. This result, together with the observation that a de novo KLF1 site formed by the –198T-to-C mutation can upregulate fetal globin (***Tate et al., 1986***; ***Wienert et al., 2017***) raises the possibility that introduction of a KLF1 binding site anywhere around the *HBG* promoter could potentially upregulate *HBG* gene expression. In contrast to our finding in EMSA, we observed only a very weak signal for the binding of KLF1 at the edited site of the *HBG* promoter by ChIP. Thus, our hypothesis, primarily on the basis of observing in vitro binding of KLF1 in EMSAs, is that the –123 and –124 mutations create a new KLF1 binding site, that is relatively weak and difficult to detect using ChIP but other hypotheses are possible. For instance, it could create a binding site for another activator. The relative proximity of this site to the BCL11A site, that begins around –117, suggests it may also directly or indirectly affect BCL11A binding. Further work needs to be done to assess these possibilities.

The current screening approaches that we used to identify the regulatory element in the proximal promoter of *HBG* is limited by several technical issues. The availability of NGG PAM sequences in the target region confines the resolution of the screening approach. The editing efficiency for ABE7.10 RA or BE3RA-FNLS is not uniform across the target regions (***Koblan et al., 2018***). The effect of transverse mutation in the target region on gene regulation is not possible as the current base editors are mainly involved in the installation of transition mutations (***Gaudelli et al., 2017***; ***Komor et al., 2016***). The bystander mutation introduced by the base editors at the target regions makes it difficult to identify functional regulatory single nucleotides responsible for the gamma-globin regulation. These limitations can be overcome by the use of several different strategies including the use of alternative base editor variants that recognize the non-canonical PAM site (***Richter et al., 2020***). In addition, recently developed hyperactive variant of base editors will improve the increasing editing efficiency at the target site with the broader editing window (***Richter et al., 2020***). The scope of this study can be further increased by the dual ABE and CBE that can mediate both conversions (A-to-G and C-to-T) simultaneously, and also by prime editing approach which can widen the range of precise conversions in the desired region (***Anzalone et al., 2019***; ***Zhang et al., 2020***).

The translational potential of genome edited HSPCs depends on long-term engraftment and repopulation ability. However, genotoxicity and cytotoxicity that can arise as a result of DSBs generated by programmable nucleases can be a limiting factor (***Cullot et al., 2019***; ***Yu et al., 2016***). A previous study in nonhuman primates observed that the *HBG* promoter editing by Cas9 resulted in *HBG2* deletion with up to 27% frequency and that cells with this deletion were under-represented after engraftment (***Humbert et al., 2019***). Base editing at the target sites of *HBG1* and *HBG2* promoter by ABE and CBE does not result in high frequency of large deletions in the intergenic region as seen with Cas9 and only showed low levels of indel formation. ABEs have an inherent advantage over CBEs as they generate desired edits (A:T to G:C) with high fidelity, whereas the latter generate unanticipated edits. In corroboration with existing findings, our results also suggest that ABE is a better base editor than CBE with respect to purity of base conversion and indel formation (***Lee et al., 2018***). Moreover, preliminary results from our study suggest that the base editing of the HSPCs by ABE8e variant with the novel site (by gRNA-11) elevated HbF to therapeutic levels in erythroid progeny. Further, our study did not observe any significant DNA and RNA off-target in the ABE and CBE edited cells. Our proof of principle study validated the various gRNAs that can elevate the HbF levels to therapeutic

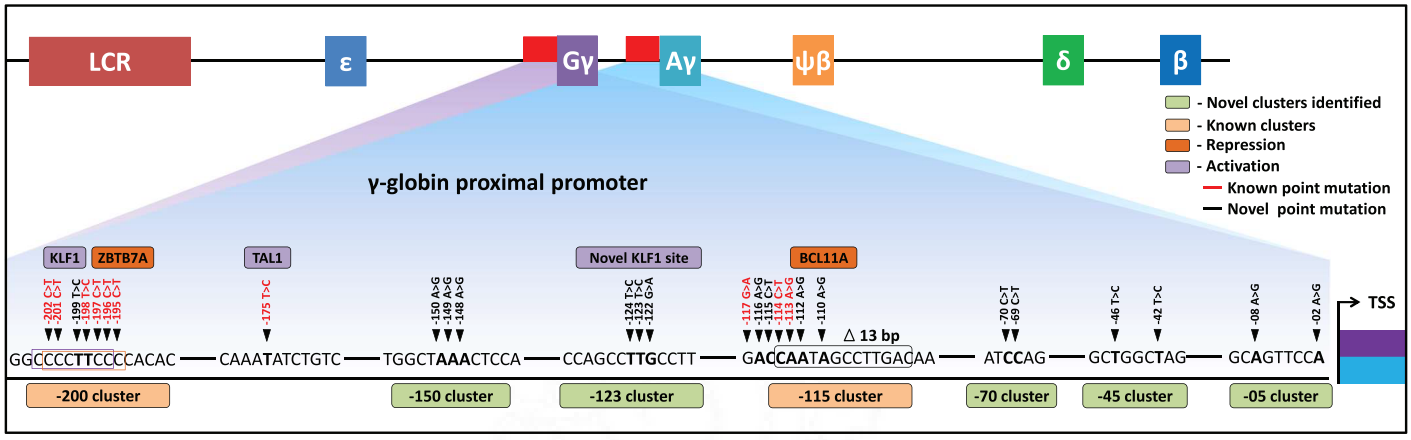


Figure 7. Schematic representation of known and identified point mutations in *HBG* promoter region that elevates fetal hemoglobin (HbF): The proximal promoter region of *HBG2* and *HBG1* is represented from transcription start site (TSS) till –205 bases.

Novel clusters identified from this study are highlighted in Sage (five clusters), and known clusters are highlighted in Melon (two clusters). Among these clusters, known base conversions are represented in black and identified hereditary persistence of fetal hemoglobin (HPFH)-like mutations are represented in red text. The novel base conversions from our study are represented in bold font. Transcriptional activators (lavender) and repressors (orange) that bind to the known clusters are also depicted in the figure.

levels laying the groundwork for potential clinical applications. This approach could address a range of beta-globin disorders avoiding the need to develop specific therapeutic products for each of them.

In summary, we have demonstrated that CRISPR base editing can be utilized to drive the expression of HbF to therapeutically relevant levels in an erythroid progenitor cell line and in HSPCs. After screening every gRNAs within the 320 bp region of the *HBG* promoter, we identified nine gRNAs that, when paired with the appropriate base editor, can introduce HPFH-like mutations without the generation of indels. We identified five novel regulatory regions for *HBG1* and *HBG2* that are required for the silencing of gamma-globin in adult erythroid cells shedding light on the molecular mechanisms behind hemoglobin switching (***Figure 7***). Our work is an exemplification of base editors in mapping gene regulatory elements in highly homologous locus and we hope base editing strategy will be among the pre-eminent therapeutic strategies for monogenetic disorders like beta-hemoglobinopathies in the future.

Materials and methods

Designing and cloning of the gRNA

The gRNAs for targeting the *HBG1* and *HBG2* promoter region were designed using SnapGene and Benchling. The gRNAs for CBE were designed using design-type 'gRNAs for base editing' in the Benchling tool; from the 43 hits, we selected 32 non-overlapping gRNAs. The gRNAs for ABE were designed manually using SnapGene software. The forward oligonucleotide consists of the gRNA sequence without PAM (20 bp) and 'CACCG' overhang at the 5' end, while the reverse oligonucleotide consists of reverse complement of gRNA without PAM (20 bp), 'AAAC' overhang at the 5' end and a 'C' added at 3' end. The synthetic complementary oligonucleotides listed in ***Supplementary file 1*** were annealed (***Ran et al., 2013b***; ***Shalem et al., 2014***) and cloned into *BsmBI* digested pLKO5.sgRNA.EFS.GFP/RFP vector (gift from Benjamin Ebert, Addgene #57822/#57823) (***Heckl et al., 2014***). The oligo-annealed products were diluted 1:200-fold, from which 6 µl was taken along with 50 ng of vector backbone and ligation reaction was set up as per the manufacturer's instruction from NEB. The ligated product was transformed into DH10B competent cells and plated in LB agar containing 100 µg/ml of ampicillin for selection (***Sambrook and Russell, 2006***). Three colonies were picked from the plate and inoculated in LB for colony PCR. Colony PCR was carried out using GoTaq Hot Start Polymerase premix (Promega) and 1 µl each of forward and reverse sequencing primers (10 picomoles) (***Supplementary file 2***) along with 1 µl of processed cells in a thermocycler (Applied Biosystems Veriti). The cyclic conditions were as follows: initial denaturation at 95°C for 10 min, 35

cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 45 s, followed by a final extension at 72°C for 7 min. After confirming the expected amplification in 1% agarose gel, second round of PCR was carried out using the 20 ng of pre-cleaned product from the first round of PCR using BigDye Terminator v3.1 Cycle Sequencing Kit as per manufacturer's protocol and given for Sanger sequencing.

Plasmid constructs

The plasmids used in this study, pLenti-FNLS-P2A-Puro (Addgene#110841-CBE) and pLenti-ABERA-P2A-Puro (Addgene#112675-ABE), were a gift from Lukas Dow (***Zafra et al., 2018***), and pMD2.G and psPAX2 (second-generation lentiviral packaging construct, Addgene #12259, 11260) were a gift from Didier Trono. The pLenti-ABE8e-puro vector was constructed by amplifying ABE8e from the TadA-8e V106W plasmid (a plasmid gifted from David Liu, Addgene#138495) (***Richter et al., 2020***) using the primers mentioned in ***Supplementary file 2***. The amplified PCR product was then cloned into pLenti-ABERA-P2A-Puro backbone after digestion with BamH1 and Nhe1 by HIFI assembly (NEB). The gRNA sequence from lentiCRISPR V2 vector (a construct gifted from Feng Zhang, Addgene#52961) was removed by digestion with EcoR1 and Kpn1 enzyme (NEB). The digested plasmid was then re-ligated with NEB Ligase after a exonuclease treatment (NEB Exonuclease 1) to generate a lentiCRISPR V2.1 vector (***Sanjana et al., 2014***). The plasmids were isolated using NucleoBond Xtra Midi EF (Macherey-Nagel) according to the manufacturer's instruction.

Cell culture

HUDEP-2 cell lines were cultured in StemSpan SFEM II (STEMCELL Technologies) supplemented with 50 ng/ml SCF (ImmunoTools), 3 U/ml EPO (Zyrop 4000 IU injection), 1× Pen-Strep (Gibco), 1 µM dexamethasone (Alfa Aesar), 1 µg/ml doxycycline (Sigma-Aldrich), and 1× L-glutamine 200 mM (Gibco) (***Kurita et al., 2013***). The cells were culture at 37°C with 5% CO_2 and were confirmed negative for mycoplasma (Universal Mycoplasma detection kit-ATCC). K562 cell line was cultured in RPMI (Roswell Park Memorial Institute media) (Hyclone) supplemented with 1× penicillin-streptomycin-glutamine (Gibco) and 10% fetal bovine serum (FBS) (Gibco). COS-7 cells and HEK 293T cells were cultured in Dulbecco's modified Eagle medium (DMEM, Gibco) supplemented with 10% (v/v) FBS and 1× Pen-Strep.

The left-over peripheral blood mononuclear cells (PBMNCs) were obtained from a healthy donor after infusion according to the clinical protocols approved by the Intuitional Review Boards of Christian Medical College, Vellore. The PBMNCs were purified by density gradient centrifugation (Lymphoprep Density Gradient Medium|STEMCELL Technologies) followed by RBC lysis. CD34+ cells were isolated from the purified PBMNCs by EasySep Human CD34 positive selection kit II (STEMCELL Technologies) and expanded in HSC expansion media as described earlier (***Genovese et al., 2014***). The isolated cells were analyzed for primitive cell surface markers (CD34+ CD133+ and CD90+) after 24 hr of expansion (***Genovese et al., 2014***).

Lentivirus production

HEK293T cells (1×10^6) were cultured in 10 cm cell culture dish (Corning). Around 80% confluency, 2.5 µg of pMD2.G (envelope plasmid), and 3.5 µg of psPAX2 (packaging plasmid) along with 4 µg (construct with gRNA) or 5 µg (construct with ABE/CBE/Cas9) of lentiviral vector were transfected using FuGENE-HD as per the manufacturer's protocol. The viral supernatants were separately collected at 48 and 72 hr; and concentrated using Lenti-X Concentrator (Takara). The concentrated pellet was resuspended in 200 µl of 1×PBS, and the aliquots were stored at –80°C.

Lentiviral transduction

The desired lentivirus (100 µl aliquot) along with 6 µg/ml polybrene (Sigma-Aldrich), and 1% HEPES 1 M buffer (Gibco) were added to HUDEP-2 or K562 cells (0.5 million cells in one well of a six-well plate) and spinfected at 800 g for 30 min at room temperature. The cells were incubated for 48 hr with lentivirus at 37°C and then incubated in fresh medium. For the stable cell line generation, the cells transduced with pLenti-FNLS-P2A-puro or pLenti-ABERA-P2A-puro or pLenti-ABE8e-puro or lentiCRISPR V2.1viral vector were then treated with 1 µg/ml puromycin (Gibco) for 10 days. In case of gRNA transduction with pLKO5.sgRNA.EFS.GFP/RFP vector, the transduced cells were analyzed by FACS for GFP/RFP expression.

In vitro transcription

The template for in vitro transcription (IVT) was prepared by linearizing ABE8e plasmid (Addgene#138495) with Pme1 (NEB) and purified using PCI (phenol-chloroform-isoamyl alcohol). IVT was carried out using T7 mScript Standard mRNA Production System (CELLSCRIPT) components by previously described method with full substitution of pseudouridine-5'-triphosphate (Jena Bioscience) for uridine (***Mahalingam et al., 2022***). The purified mRNA was stored as aliquots (5 µg/vial) in –80°C.

Electroporation of CD34+ cells

CD34+ cells were expanded in HSC expansion media for 48 hr. Around 1 million of CD34+ cells were pelleted then resuspended in 19 µl MaxCyte buffer (Hyclone) along with 5 µg ABE8e mRNA (5 µl) and 100 pmole desired gRNA (1 µl) (Synthego) (target information in ***Supplementary files 1 and 2***). The resuspended cells were loaded into one well of OC25 × 3 Maxcyte cuvette and electroporated with program 'HSC-3'. After electroporation, the content was transferred to single well of 12-well plate (Corning) and allowed to recover for 20 min in the incubator (5% CO_2, 37°C). To the recovered cells, 1 ml of HSC expansion media was added and then expanded for 48 hr before performing any further experiments.

Erythroid differentiation

For the erythroid differentiation of HUDEP-2 cells, we followed previously established protocol with slight modification (***Trakarnsanga et al., 2017***). After 8 days of expansion, around 1 million of edited cells were seeded in 65 mm cell culture dish (Eppendorf) with 5 ml of differentiation media consisting of IMDM glutamax (Gibco), 3% AB serum (MP Biomedicals), 2% FBS, 0.1% insulin solution human (Sigma-Aldrich), 3 U/ml Heparin sodium salt (MP Biomedicals), 200 µg/ml Holo Transferrin (BBI Solutions), 3 U/ml EPO, 10 µg/ml SCF, 1 ng/ml IL3 (Immuno Tools), 1× Pen-Strep, and 1 µg/ml doxycycline. Erythroid differentiation was carried out in 10 cm dish with regular media change (on days 3 and 6) up to the end of differentiation (for 9 days). On day 6, these cells were cultured in erythroid differentiation medium with 500 µg/ml of holotransferrin and devoid of doxycycline.

For erythroid differentiation of CD34+ cells, HSPCs were cultured in a three-phase liquid culture system and subjected to enucleation analysis as previously described (***Psatha et al., 2018***). The erythroid differentiation pattern was evaluated in the erythroblast obtained from HUDEP-2 cells (on day 9) and CD34+ cells (on day 21) by FACS analysis of CD235a and CD71 markers.

Analysis of base editing efficiency

Genomic DNA was isolated from the edited samples using DNA isolation kit (NucleoSpin Blood – Macherey-Nagel). For Sanger sequencing, the targets were PCR amplified using GoTaq Hot Start Polymerase premix (Promega), the primers used are listed in ***Supplementary files 1 and 2***. For NGS, the targets were PCR amplified (the primers listed in ***Supplementary files 1 and 2***) using GXL premix (Takara Bio) and sequenced using MiSeq System (Illumina). The library preparation and sequencing was carried out as per previously described protocol (***Corn, 2017***). The Fastq files obtained were analyzed for base editing using CRISPResso-2 (***Clement et al., 2019***). The data obtained from Sanger sequencing were used to analyze indels and the base editing efficiency by tools like Inference of CRISPR Edits (ICE) (Synthego) and EditR, respectively (***Hsiau et al., 2018***; ***Kluesner et al., 2018***).

For characterization of editing in individual *HBG1* and *HBG2* promoter, NGS 4F and NGS 2R primers (***Supplementary files 1 and 2***) were used to amplify *HBG* promoter. After NGS , the Fastq file obtained were aligned to *HBG1* and *HBG2* sequence using Bowtie2 based on nucleotide variation at −307, –317, and –324. The aligned reads were visualized using IGV (***Robinson, 2012***; ***Langmead and Salzberg, 2013***) and the editing efficiency was computed individually for both the genes ((edited reads/total reads) ×100).

Real-time PCR

Total RNA from the edited cells were isolated using the NucleoSpin RNA kit (Macherey-Nagel) and reverse transcribed using cDNA Synthesis Kit (iScript Bio-Rad). The relative expression (ΔΔCT) of *HBB*, *HBA*, and *HBG* genes was determined using the respective primers (***Supplementary file 2***) by qRT-PCR using SsoFast EvaGreen Supermixes (Bio-Rad) in QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). The qRT-PCR mixture (10 µl) contains 1 µl each of respective forward and reverse primer

(5 µM), 5 µl of SYBR green master mix, 2 µl of H_2O, and 1 µl of 5-fold diluted cDNA template. *GAPDH* was used as an internal control gene to normalize the data for ΔΔCT (relative expression analysis). The cycling condition was performed as per the manufacturer's protocol (Bio-Rad). A dissociation curve analysis was carried out to ensure there is no unspecific amplification.

The VCN was assessed in genomic DNA isolated from the transduced samples using qRT-PCR as previously described with a few modification (***Barczak et al., 2015***). Primers targeting U6 promoter (for gRNA integration), *Cas9* gene (for Cas9 and base editors' integration), and WPRE (for gRNA and Cas9 variant integration together) were used. Exon 2 of *HBB* gene was used as a single copy gene-specific reference. The primers used are listed in ***Supplementary file 2***. pLKO5.sgRNA. EFS.GFP (Plasmid #57822, Addgene) and an inhouse plasmid carrying *HBB* CDS (details not provided) were used as standards.

HbF intracellular staining

To evaluate the frequency of HbF positive cells, the cells were fixed, permeabilized, and intracellular staining was performed using Fetal Hemoglobin Monoclonal Antibody (HBF-1), APC (Invitrogen) as previously described (***Canver et al., 2015***). The stained cells were analyzed by FACS (BD FACSAria III Cell Sorter or CytoFLEX LX Flow Cytometer – BC) to measure the number of HbF positive cells.

Hemoglobin detection by HPLC

The differentiated cells were collected and washed with 1× PBS and resuspended in 1100 µl cold ddH_2O. The cells were sonicated for 30 s with 50% Amp in ice and centrifuged at 14,000 rpm for 15 min at 4°C. The supernatant (1000 µl) was analyzed for hemoglobin variants by VARIANT II Hemoglobin Testing System (Bio-Rad). The hemoglobin percentages were calculated by the Bio-Rad's Clinical Data Management (CDMTM) Software. Reverse phase HPLC (Shimadzu Corporation-Phenomenex) (***Loucari et al., 2018***) was performed in remaining 100 µl of the supernatant for the analysis of individual globin chains expression . The ratio of gamma (A and G gamma)/beta-like (gamma, beta, and delta) globin was calculated and represented in percentage.

Validation of 4.9 kb large deletion

To quantify the large deletion in *HBG* promoter region, qPCR was carried as previously reported (***Li et al., 2018***) (using primers from ***Supplementary file 2***). To verify the effect of larger deletion on gamma-globin expression, the globin chain analysis was carried out using RP HPLC in the differentiated erythroid cells. The A gamma- and G gamma-globin chain percentage obtained from each sample were normalized with control.

COS-7 cell transfections and nuclear extraction

COS-7 cells were transfected with 5 µg of mammalian expression plasmids pcDNA3-empty (Invitrogen) or pSG5/mEKLF-Mouse (***Miller and Bieker, 1993***) using FuGENE 6 (Promega) in 10 cm culture dish, according to the manufacturer's instructions. Transfected cells were incubated at 37°C for 48 hr before harvest. Nuclear extractions were performed as previously described (***Andrews and Faller, 1991***).

Electrophoretic mobility shift assay

Oligonucleotides used in radiolabelled probes are listed in ***Supplementary file 2***. The sense strand for each probe was labelled with P-32 from γ-^{32}P ATP (Perkin Elmer) using T4 PNK (NEB), before annealing the antisense strand by slow cooling from 100°C to room temperature. The annealed probes were purified using quick spin columns for radiolabelled DNA purification (Roche). Plasmids were overexpressed and harvested from COS-7 cells, and 'empty' extract without the target protein was used to aid identification of background bands caused by endogenous protein binding. Antibody for KLF1 was used as indicated to identify the protein on the gel (***Crossley et al., 1996***). Complexed samples were loaded on 6% native polyacrylamide gel in TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA). Electrophoresis was performed at 4°C and 250 V for 1 hr and 40 min, and then vacuum dried before exposing a FUJIFILM BAS Cassette2 phosphor screen overnight. Imaging was performed on a GE Typhoon FLA 9500 fluorescent image analyzer.

ChIP qPCR

Each immunoprecipitation was performed using 5 × 10^7 cells of wild type and edited HUDEP-2 cells before differentiation. Cells were cross-linked in 1% formaldehyde solution and incubated at room temperature for 10 min before the reaction was quenched by addition of glycine to a final concentration of 125 mM. Cross-linked cells were lysed and sonicated for 10 cycles of 30 s with 30 s intermissions at 4°C to obtain chromatin fragments of approximately 200–300 bp. Immunoprecipitations were performed using 100 µl of Dynabeads Protein G (ThermoFisher Scientific) complexed to 15 µg of KLF1 antibody (OriGene, #TA305808) or normal rabbit IgG (Cell Signaling Technology #2729S) at 4°C overnight. Magnetic beads were separated and washed thoroughly before elution and cross-linking was reversed by incubation at 65°C overnight. DNA was then purified and quantified within reference to whole cell extract on a ViiA 7 Real-Time PCR System using SYBR green reagents and the ΔΔCt method for specific targets (***Supplementary file 2***).

RNA sequencing and analysis

Total RNA extracted using NucleoSpin RNA kit (Macherey-Nagel) was quality assessed by Agilent 2100 Bioanalyzer (Agilent Technologies). From 1 µg of total RNA polyadenylated transcripts was purified using oligo-dT beads (TruSeq RNA Sample Preparation Kit, Illumina). Fragmentation was carried out in the presence of divalent cation followed by reverse transcription using Superscript II Reverse Transcriptase kit (Life Technologies). Following cDNA purification by Ampure XP SPRI beads (Beckman Coulter) Illumina adapter ligation and amplification were carried out. The quantification and the quality were assessed by NanoDrop spectrophotometer (Thermo Scientific) and Bioanalyzer (Agilent Technologies), respectively. Libraries were sequenced by using Illumina NovoSeq 6000 platform as 150 bp paired-end reads. Fastq files were generated with bcl2fastq and then trimmed to remove low-quality bases, adapter seq, and unpaired sequence using TrimGalore. *Homo sapiens* genome assembly GRCh38 was used as a reference to align the trimmed reads. NFCore RNA Seq pipeline was used to resolve the expressed transcripts quantitatively and qualitatively (***Ewels et al., 2020***). The files are accessible through the GEO Series accession number GSE192801.

Transcriptome analysis was carried out in wild type HUDEP-2, ABE, and CBE stable cells with or without gRNA-2 and -11 in duplicate. The transcript was counted from the sorted bam files by the aligner mentioned above. Interactive Gene Expression Analysis Kit (iGEAK) RNA-seq v1.0 a R and JavaScript-based tool was used to normalize gene expression levels and perform differential expression analysis (***Choi and Ratner, 2019***).

Off-target analysis

Cas-OFFinder was used to find the Cas-dependent DNA off-target, up to three mismatches were allowed in selecting targets (target information in ***Supplementary file 3***). The targets were amplified and sequenced using Illumina MiSeq platform (using primers mentioned in ***Supplementary file 2***). CRISPResso2 was used to align the reads, only high-quality reads were used for this analysis (q = 30). REDItools v2 was used to calculate the transcriptome-wide A-to-I and C-to-U conversion in ABE and CBE edited samples. Except the respective nucleotide (A for ABE and C for CBE), all nucleotides were removed from the analysis. Read coverage and read quality criteria were followed as described earlier (***Koblan et al., 2021***). The frequency of A converted to I/N and C converted to U/N was calculated by dividing the total number of converted nucleotides by the respective nucleotides after filtering (A-to-(I or N)/A*100 or C-to-(U or N)/C*100). The experiment was carried out as two biological replicates.

Statistical analysis

The statistical tests were performed using GraphPad Prism 8.1. Since all the data were normally distributed, unpaired two-sided t-test or one-way ANOVA was used as appropriate. In all the tests, $p < 0.05$ was considered statistically significant. Linear regression was carried out to find out if any correlation exists between two variables. Also, to find the relationship between the samples, Pearson correlation was performed. Principal component analysis (PCA) was performed using R statistical package.

Acknowledgements

The research reported in this work was supported by NAHD grant: BT/PR17316/MED/31/326/2015 (Department of Biotechnology, New Delhi, India), EMR grant: EMR/2017/004363 (Science and Engineering Research Board [SERB], New Delhi, India), Indo-US GETin Fellowship_2018_066 (Indo-US Science & Technology Forum [IUSSTF]), and DBT grant: BT/PR38392/GET/119/301/2020. We sincerely acknowledge CSCR (a unit of inStem, CMC Campus, Vellore, India) for providing the startup funds. NSR and AG is supported by Senior Research Fellowship from Council of Scientific & Industrial Research India. VR is supported by Senior Research Fellowship DBT India. BW is supported by an Early Career Research Fellowship, and HWB and MC were supported by a grant from the National Health and Medical Research Council Australia. HWB is additionally supported by an Australian Government Research Training Program Scholarship. SM is supported by gene editing task force (DBT); Grant No# BT/PR25841/GET/119/162/2017. We thank Mrs Sumithra and Mr Neelagandan at the Department of Hematology, CMC, for help with HPLC variants; Keerthivasan. RC, IISER Mohali, and Ashis Kumar S, CSCR, for bioinformatics. Also, we like to acknowledge the CSCR core facility for supporting us with all the required instrumentations.

Additional information

Funding

Funder	Grant reference number	Author
Ministry of Science and Technology	BT/PR17316/MED/31/326/2015	Kumarasamypet M Mohankumar
Science and Engineering Research Board	EMR/2017/004363	Kumarasamypet Murugesan Mohankumar
Indo-US Science and Technology Forum	Indo-U.S. GETin Fellowship_2018_066	Kumarasamypet Murugesan Mohankumar
Ministry of Science and Technology	BT/PR38392/GET/119/301/2020	Kumarasamypet M Mohankumar
Council of Scientific and Industrial Research, India	Senior Research Fellow	Nithin Sam Ravi Anila George
Ministry of Science and Technology	Senior Research Fellow	Vignesh Rajendiran
National Health and Medical Research Council	Early Career Research Fellowship	Beeke Wienert
Ministry of Science and Technology	BT/PR25841/GET/119/162/2017	Srujan Marepally
National Health and Medical Research Council	National Health and Medical Research Council (NHMRC)	Henry William Bell
National Health and Medical Research Council	Grant	Merlin Crossley
National Health and Medical Research Council		Henry William Bell

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Nithin Sam Ravi, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing; Beeke Wienert, Resources, Visualization, Writing – review and editing; Stacia K Wyman, Methodology, Resources, Software; Henry William Bell, Jonathan T Vu, Aswin Anand Pai, Poonkuzhali Balasubramanian, Methodology, Resources; Anila George, Formal analysis, Writing – review and editing, Methodology; Gokulnath Mahalingam, Kirti Prasad, Bhanu Prasad Bandlamudi, Nivedhitha Devaraju, Vignesh

Rajendiran, Nazar Syedbasha, Methodology; Yukio Nakamura, Ryo Kurita, Resources; Muthuraman Narayanasamy, Writing – review and editing; Saravanabhavan Thangavel, Srujan Marepally, Resources, Writing – review and editing; Shaji R Velayudhan, Mark A DeWitt, Methodology, Resources, Writing – review and editing; Alok Srivastava, Project administration, Resources, Writing – review and editing; Merlin Crossley, Investigation, Methodology, Resources, Visualization, Writing – review and editing; Jacob E Corn, Methodology, Resources, Visualization, Writing – review and editing; Kumarasamypet M Mohankumar, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

Author ORCIDs
Nithin Sam Ravi http://orcid.org/0000-0002-8063-6935
Henry William Bell http://orcid.org/0000-0003-4677-8208
Anila George http://orcid.org/0000-0002-6016-976X
Jonathan T Vu http://orcid.org/0000-0002-4950-7967
Jacob E Corn http://orcid.org/0000-0002-7798-5309
Kumarasamypet M Mohankumar http://orcid.org/0000-0001-9407-1800

Ethics
The left-over peripheral blood mononuclear cells (PBMNC) were obtained from a healthy donor after infusion according to the clinical protocols approved by the Instituitional Review Boards of Christian Medical College, Vellore. IRB Min. No. 12309 (OTHER) dated 30. 10.2019.

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.65421.sa1
Author response https://doi.org/10.7554/eLife.65421.sa2

Additional files

Supplementary files
- Supplementary file 1. The guide RNAs (gRNAs) used in this study to screen the *HBG* promoter region and their respective primer for sequencing.
- Supplementary file 2. All the PCR, qRT-PCR primers and probes used in this study.
- Supplementary file 3. The targets analyzed for DNA off-target.
- Transparent reporting form

Data availability
The transcriptome data have been deposited in GEO under accession code GSE192801 All the raw data from this study have been deposited in Dyrad (https://doi.org/10.5061/dryad.bzkh1897h).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Ravi N, Wyman SK, Mohankumar KM	2022	Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192801	NCBI Gene Expression Omnibus, GSE192801
Mohankumar KM	2022	Data from: Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin	http://dx.doi.org/10.5061/dryad.bzkh1897h	Dryad Digital Repository, 10.5061/dryad.bzkh1897h

References
Andrews NC, Faller DV. 1991. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Research* **19**:2499. DOI: https://doi.org/10.1093/nar/19.9.2499, PMID: 2041787

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**:149–157. DOI: https://doi.org/10.1038/s41586-019-1711-4, PMID: 31634902

Anzalone AV, Koblan LW, Liu DR. 2020. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology* **38**:824–844. DOI: https://doi.org/10.1038/s41587-020-0561-9, PMID: 32572269

Arbab M, Shen MW, Mok B, Wilson C, Matuszek Ż, Cassa CA, Liu DR. 2020. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* **182**:463-480.. DOI: https://doi.org/10.1016/j.cell.2020.05.037, PMID: 32533916

Barczak W, Suchorska W, Rubiś B, Kulcenty K. 2015. Universal real-time PCR-based assay for lentiviral titration. *Molecular Biotechnology* **57**:195–200. DOI: https://doi.org/10.1007/s12033-014-9815-4, PMID: 25370825

Bauer DE, Orkin SH. 2012. Update on fetal hemoglobin gene regulation in hemoglobinopathies. *Current Opinion in Pediatrics* **23**:617–632. DOI: https://doi.org/10.1097/MOP.0b013e3283420fd0.Update

Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O, Chen DD, Schupp PG, Vinjamur DS, Garcia SP, Luc S, Kurita R, Nakamura Y, Fujiwara Y, Maeda T, Yuan GC, Zhang F, Orkin SH, Bauer DE. 2015. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**:192–197. DOI: https://doi.org/10.1038/nature15521, PMID: 26375006

Cavazzana M, Antoniani C, Miccio A. 2017. Gene Therapy for β-Hemoglobinopathies. *Molecular Therapy* **25**:1142–1154. DOI: https://doi.org/10.1016/j.ymthe.2017.03.024, PMID: 28377044

Choi K, Ratner N. 2019. iGEAK: an interactive gene expression analysis kit for seamless workflow using the R/shiny platform. *BMC Genomics* **20**:1–7. DOI: https://doi.org/10.1186/s12864-019-5548-x, PMID: 30841853

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L. 2019. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* **37**:224–226. DOI: https://doi.org/10.1038/s41587-019-0032-3, PMID: 30809026

Corn JE. 2017. Preparation of PCR amplicons from edited cells for deep sequencing. Protocols. https://www.protocols.io/view/preparation-of-pcr-amplicons-from-edited-cells-for-6ruhd6w

Crossley M, Whitelaw E, Perkins A, Williams G, Fujiwara Y, Orkin SH. 1996. Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells. *Molecular and Cellular Biology* **16**:1695–1705. DOI: https://doi.org/10.1128/MCB.16.4.1695, PMID: 8657145

Cullot G, Boutin J, Toutain J, Prat F, Pennamen P, Rooryck C, Teichmann M, Rousseau E, Lamrissi-Garcia I, Guyonnet-Duperat V, Bibeyran A, Lalanne M, Prouzet-Mauléon V, Turcq B, Ged C, Blouin JM, Richard E, Dabernat S, Moreau-Gaudry F, Bedel A. 2019. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature Communications* **10**:1136. DOI: https://doi.org/10.1038/s41467-019-09006-2, PMID: 30850590

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* **38**:276–278. DOI: https://doi.org/10.1038/s41587-020-0439-x

Fischer KD, Nowock J. 1990. The T----C substitution at -198 of the A gamma-globin gene associated with the British form of HPFH generates overlapping recognition sites for two DNA-binding proteins. *Nucleic Acids Research* **18**:5685–5693. DOI: https://doi.org/10.1093/nar/18.19.5685, PMID: 1699206

Fucharoen S, Shimizu K, Fukumaki Y. 1990. A novel C-T transition within the distal CCAAT motif of the G gamma-globin gene in the Japanese HPFH: implication of factor binding in elevated fetal globin expression. *Nucleic Acids Research* **18**:5245–5253. DOI: https://doi.org/10.1093/nar/18.17.5245, PMID: 1698280

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**:464–471. DOI: https://doi.org/10.1038/nature24644, PMID: 29160308

Genovese P, Schiroli G, Escobar G, Tomaso TD, Firrito C, Calabria A, Moi D, Mazzieri R, Bonini C, Holmes MC, Gregory PD, van der Burg M, Gentner B, Montini E, Lombardo A, Naldini L. 2014. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**:235–240. DOI: https://doi.org/10.1038/nature13420, PMID: 24870228

Gilman JG, Mishima N, Wen XJ, Stoming TA, Lobel J, Huisman TH. 1988. Distal CCAAT box deletion in the A gamma globin gene of two black adolescents with elevated fetal A gamma globin. *Nucleic Acids Research* **16**:10635–10642. DOI: https://doi.org/10.1093/nar/16.22.10635, PMID: 2462713

Giudice CL. 2022. REDItools2. GitHub. https://github.com/BioinfoUNIBA/REDItools2

Gluzman Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**:175–182. DOI: https://doi.org/10.1016/0092-8674(81)90282-8, PMID: 6260373

Heckl D, Kowalczyk MS, Yudovich D, Belizaire R, Puram RV, McConkey ME, Thielke A, Aster JC, Regev A, Ebert BL. 2014. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature Biotechnology* **32**:941–946. DOI: https://doi.org/10.1038/nbt.2951, PMID: 24952903

Hsiau T, Conant D, Rossi N, Maures T, Waite K, Yang J, Joshi S, Kelso R, Holden K, Enzmann BL, Stoner R. 2018. Inference of CRISPR Edits from Sanger Trace Data. [bioRxiv]. DOI: https://doi.org/10.1101/251082

Humbert O, Radtke S, Samuelson C, Carrillo RR, Perez AM, Reddy SS, Lux C, Pattabhi S, Schefter LE, Negre O, Lee CM, Bao G, Adair JE, Peterson CW, Rawlings DJ, Scharenberg AM, Kiem H-P. 2019. Therapeutically relevant engraftment of a CRISPR-Cas9-edited HSC-enriched population with HbF reactivation in nonhuman primates. *Science Translational Medicine* **11**:1–14. DOI: https://doi.org/10.1126/scitranslmed.aaw3768, PMID: 31366580

Ikuta T, Papayannopoulou T, Stamatoyannopoulos G, Kan YW. 1996. Globin Gene Switching. *Journal of Biological Chemistry* **271**:14082–14091. DOI: https://doi.org/10.1074/jbc.271.24.14082, PMID: 8662960

Jacob GF, Raper AB. 1958. Hereditary persistence of foetal haemoglobin production, and its interaction with the sickle-cell trait. *British Journal of Haematology* **4**:138–149. DOI: https://doi.org/10.1111/j.1365-2141.1958.tb03844.x, PMID: 13536251

Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahante JE, Webber BR, Moriarity BS. 2018. EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal* **1**:239–250. DOI: https://doi.org/10.1089/crispr.2018.0014, PMID: 31021262

Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A, Liu DR. 2018. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology* **36**:843–846. DOI: https://doi.org/10.1038/nbt.4172, PMID: 29813047

Koblan LW, Erdos MR, Wilson C, Cabral WA, Levy JM, Xiong ZM, Tavarez UL, Davison LM, Gete YG, Mao X, Newby GA, Doherty SP, Narisu N, Sheng Q, Krilow C, Lin CY, Gordon LB, Cao K, Collins FS, Brown JD, et al. 2021. In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature* **589**:608–614. DOI: https://doi.org/10.1038/s41586-020-03086-7, PMID: 33408413

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**:420–424. DOI: https://doi.org/10.1038/nature17946, PMID: 27096365

Kurita R, Suda N, Sudo K, Miharada K, Hiroyama T, Miyoshi H, Tani K, Nakamura Y. 2013. Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLOS ONE* **8**:e0059890. DOI: https://doi.org/10.1371/journal.pone.0059890

Kuzminov A. 2001. Single-strand interruptions in replicating chromosomes cause double-strand breaks. *PNAS* **98**:8241–8246. DOI: https://doi.org/10.1073/pnas.131009198, PMID: 11459959

Langmead B, Salzberg S. 2013. Fast gapped-read alignment with Bowtie 2 Ben. *Nature Methods* **9**:357–359. DOI: https://doi.org/10.1038/nmeth.1923.Fast

Lee HK, Willi M, Miller SM, Kim S, Liu C, Liu DR, Hennighausen L. 2018. Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nature Communications* **9**:7–12. DOI: https://doi.org/10.1038/s41467-018-07322-7, PMID: 30442934

Li C, Psatha N, Sova P, Gil S, Wang H, Kim J, Kulkarni C, Valensisi C, Hawkins RD, Stamatoyannopoulos G, Lieber A. 2018. Reactivation of γ-globin in adult β-YAC mice after ex vivo and in vivo hematopoietic stem cell genome editing. *Blood* **131**:2915–2928. DOI: https://doi.org/10.1182/blood-2018-03-838540, PMID: 29789357

Liu N, Zhu Q, Xu J, Bulyk ML, Orkin SH, Liu N, Zhu Q, Hong J, Kim W, Sher F. 2018. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch Article Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. *Cell* **173**:1–13. DOI: https://doi.org/10.1016/j.cell.2018.03.016

Loucari CC, Patsali P, van Dijk TB, Stephanou C, Papasavva P, Zanti M, Kurita R, Nakamura Y, Christou S, Sitarou M, Philipsen S, Lederer CW, Kleanthous M. 2018. Rapid and Sensitive Assessment of Globin Chains for Gene and Cell Therapy of Hemoglobinopathies. *Human Gene Therapy Methods* **29**:60–74. DOI: https://doi.org/10.1089/hgtb.2017.190, PMID: 29325430

Mahalingam G, Mohan A, Arjunan P, Dhyani AK. 2022. Using Lipid Nanoparticles for the Delivery of Chemically Modified mRNA into Mammalian Cells. *JoVE* **2**:e62407. DOI: https://doi.org/10.3791/62407

Martyn GE, Wienert B, Yang L, Shah M, Norton LJ, Burdach J, Kurita R, Nakamura Y, Pearson RCM, Funnell APW, Quinlan KGR, Crossley M. 2018. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nature Genetics* **50**:498–503. DOI: https://doi.org/10.1038/s41588-018-0085-0, PMID: 29610478

Martyn GE, Wienert B, Kurita R, Nakamura Y, Quinlan KGR, Crossley M. 2019. A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. *Blood* **133**:852–856. DOI: https://doi.org/10.1182/blood-2018-07-863951, PMID: 30617196

Métais J-Y, Doerfler PA, Mayuranathan T, Bauer DE, Fowler SC, Hsieh MM, Katta V, Keriwala S, Lazzarotto CR, Luk K, Neel MD, Perry SS, Peters ST, Porter SN, Ryu BY, Sharma A, Shea D, Tisdale JF, Uchida N, Wolfe SA, et al. 2019. Genome editing of HBG1 and HBG2 to induce fetal hemoglobin. *Blood Advances* **3**:3379–3392. DOI: https://doi.org/10.1182/bloodadvances.2019000820, PMID: 31698466

Miller IJ, Bieker JJ. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Molecular and Cellular Biology* **13**:2776–2786. DOI: https://doi.org/10.1128/mcb.13.5.2776-2786.1993, PMID: 7682653

Motum PI, Deng ZM, Huong L, Trent RJ. 1994. The Australian type of nondeletional G gamma-HPFH has a C-->G substitution at nucleotide -114 of the G gamma gene. *British Journal of Haematology* **86**:219–221. DOI: https://doi.org/10.1111/j.1365-2141.1994.tb03284.x, PMID: 7516698

Psatha N, Reik A, Phelps S, Zhou Y, Dalas D, Yannaki E, Levasseur DN, Urnov FD, Holmes MC, Papayannopoulou T. 2018. Disruption of the BCL11A Erythroid Enhancer Reactivates Fetal Hemoglobin in Erythroid Cells of Patients with β-Thalassemia Major. *Molecular Therapy. Methods & Clinical Development* **10**:313–326. DOI: https://doi.org/10.1016/j.omtm.2018.08.003, PMID: 30182035

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**:1380–1389. DOI: https://doi.org/10.1016/j.cell.2013.08.021, PMID: 23992846

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013b. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**:2281–2308. DOI: https://doi.org/10.1038/nprot.2013.143, PMID: 24157548

Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, Doudna JA, Liu DR. 2020. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology* **38**:883–891. DOI: https://doi.org/10.1038/s41587-020-0453-z, PMID: 32433547

Robinson JT. 2012. Integrative Genomics Viewer. *Nature Biotechnology* **29**:24–26. DOI: https://doi.org/10.1038/nbt.1754.Integrative

Sambrook J, Russell DW. 2006. Preparation and Transformation of Competent *E. coli* Using Calcium Chloride. *CSH Protocols* **2006**:pdb.prot3932. DOI: https://doi.org/10.1101/pdb.prot3932, PMID: 22485377

Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening HHS Public Access Supplementary Material. *Nature Methods* **11**:783–784. DOI: https://doi.org/10.1038/nmeth.3047.Improved

Schiml S, Fauser F, Puchta H. 2016. Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. *PNAS* **113**:7266–7271. DOI: https://doi.org/10.1073/pnas.1603823113, PMID: 27307441

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)* **343**:84–87. DOI: https://doi.org/10.1126/science.1247005, PMID: 24336571

Shariati SA, Dominguez A, Xie S, Wernig M, Qi LS, Skotheim JM. 2019. Reversible Disruption of Specific Transcription Factor-DNA Interactions Using CRISPR/Cas9. *Molecular Cell* **74**:622-633.. DOI: https://doi.org/10.1016/j.molcel.2019.04.011, PMID: 31051141

Stoming T, Stoming G, Lanclos K, Fei Y, Altay C, Kutlar F, Huisman T. 1989. An A gamma type of nondeletional hereditary persistence of fetal hemoglobin with a T----C mutation at position -175 to the cap site of the A gamma globin gene. *Blood* **73**:329–333. DOI: https://doi.org/10.1182/blood.V73.1.329.329

Tallack MR, Whitington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, Perkins AC. 2010. A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Research* **20**:1052–1063. DOI: https://doi.org/10.1101/gr.106575.110, PMID: 20508144

Tate VE, Wood WG, Weatherall DJ. 1986. The British form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to an S1 hypersensitive site 5′ to the A gamma globin gene. *Blood* **68**:1389–1393. DOI: https://doi.org/10.1182/blood.V68.6.1389.bloodjournal6861389, PMID: 2430647

Thein SL. 2018. Molecular basis of β thalassemia and potential therapeutic targets. *Blood Cells, Molecules & Diseases* **70**:54–65. DOI: https://doi.org/10.1016/j.bcmd.2017.06.001, PMID: 28651846

Trakarnsanga K, Griffiths RE, Wilson MC, Blair A, Satchwell TJ, Meinders M, Cogan N, Kupzig S, Kurita R, Nakamura Y, Toye AM, Anstee DJ, Frayne J. 2017. An immortalized adult human erythroid line facilitates sustainable and scalable generation of functional red cells. *Nature Communications* **8**:14750. DOI: https://doi.org/10.1038/ncomms14750, PMID: 28290447

Traxler EA, Yao Y, Wang Y-D, Woodard KJ, Kurita R, Nakamura Y, Hughes JR, Hardison RC, Blobel GA, Li C, Weiss MJ. 2016. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nature Medicine* **22**:987–990. DOI: https://doi.org/10.1038/nm.4170, PMID: 27525524

Webber BR, Lonetree C-L, Kluesner MG, Johnson MJ, Pomeroy EJ, Diers MD, Lahr WS, Draper GM, Slipek NJ, Smeester BA, Lovendahl KN, McElroy AN, Gordon WR, Osborn MJ, Moriarity BS. 2019. Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nature Communications* **10**:5222. DOI: https://doi.org/10.1038/s41467-019-13007-6, PMID: 31745080

Wienert B, Funnell APW, Norton LJ, Pearson RCM, Wilkinson-White LE, Lester K, Vadolas J, Porteus MH, Matthews JM, Quinlan KGR, Crossley M. 2015. Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nature Communications* **6**:1–8. DOI: https://doi.org/10.1038/ncomms8085, PMID: 25971621

Wienert B, Martyn GE, Kurita R, Nakamura Y, Quinlan KGR, Crossley M. 2017. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood* **130**:803–807. DOI: https://doi.org/10.1182/blood-2017-02-767400, PMID: 28659276

Wienert B, Martyn GE, Funnell APW, Quinlan KGR, Crossley M. 2018. Wake-up Sleepy Gene: Reactivating Fetal Globin for β-Hemoglobinopathies. *Trends in Genetics* **34**:927–940. DOI: https://doi.org/10.1016/j.tig.2018.09.004, PMID: 30287096

Wimberly H, Shee C, Thornton PC, Sivaramakrishnan P, Rosenberg SM, Hastings PJ. 2013. R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nature Communications* **4**:1–10. DOI: https://doi.org/10.1038/ncomms3115, PMID: 23828459

Yang Y, Xu Z, He C, Zhang B, Shi Y, Li F. 2019. Structural insights into the recognition of γ-globin gene promoter by BCL11A. *Cell Research* **29**:960–963. DOI: https://doi.org/10.1038/s41422-019-0221-0, PMID: 31467406

Yu K-R, Corat MAF, Metais J-Y, Dunbar CE. 2016. 564. The Cytotoxic Effect of RNA-Guided Endonuclease Cas9 on Human Hematopoietic Stem and Progenitor Cells (HSPCs). *Molecular Therapy* **24**:S225–S226. DOI: https://doi.org/10.1016/S1525-0016(16)33372-X

Zafra MP, Schatoff EM, Katti A, Foronda M, Breinig M, Schweitzer AY, Simon A, Han T, Goswami S, Montgomery E, Thibado J, Kastenhuber ER, Sánchez-Rivera FJ, Shi J, Vakoc CR, Lowe SW, Tschaharganeh DF, Dow LE. 2018. Optimized base editors enable efficient editing in cells, organoids and mice. *Nature Biotechnology* **36**:888–893. DOI: https://doi.org/10.1038/nbt.4194, PMID: 29969439

Zertal-Zidani S, Merghoub T, Ducrocq R, Gerard N, Satta D, Krishnamoorthy R. 1999. A novel C-->A transversion within the distal CCAAT motif of the Ggamma-globin gene in the Algerian Ggammabeta+-hereditary persistence of fetal hemoglobin. *Hemoglobin* **23**:159–169. DOI: https://doi.org/10.3109/03630269908996160, PMID: 10335983

Zhang X, Zhu B, Chen L, Xie L, Yu W, Wang Y, Li L, Yin S, Yang L, Hu H, Han H, Li Y, Wang L, Chen G, Ma X, Geng H, Huang W, Pang X, Yang Z, Wu Y, et al. 2020. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nature Biotechnology* **38**:856–860. DOI: https://doi.org/10.1038/s41587-020-0527-y, PMID: 32483363

Appendix 1

Appendix 1—key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Genetic reagent (*Homo sapiens*)	*GRCh38*	GenBank	883148	
Strain, strain background (*Escherichia coli*)	DH10B	ECOS, Yeastern Biotech	CAT # FYE507-10VL	
Recombinant DNA reagent	pLKO5.sgRNA. EFS.GFP	Addgene	Addgene_57822; RRID:Addgene_57822	
Recombinant DNA reagent	pLKO5.sgRNA. EFS.RFP	Addgene	Addgene_57823; RRID:Addgene_57823	
Recombinant DNA reagent	pLenti-FNLS-P2A-Puro	Addgene	Addgene_110841; RRID:Addgene_110841	
Recombinant DNA reagent	pLenti-ABERA-P2A-Puro	Addgene	Addgene_112675; RRID:Addgene_112675	
Recombinant DNA reagent	pMD2.G	Addgene	Addgene_12259; RRID:Addgene_12259	
Recombinant DNA reagent	psPAX2	Addgene	Addgene_12260; RRID:Addgene_12260	
Recombinant DNA reagent	TadA-8e V106W	Addgene	Addgene_138495; RRID:Addgene_138495	
Recombinant DNA reagent	lentiCRISPR V2	Addgene	Addgene_52961; RRID:Addgene_52961	
Recombinant DNA reagent	lentiCRISPRV2.1	This study		Cas9 expressing lentiviral plasmid without gRNA scaffold
Recombinant DNA reagent	pLenti-ABE8e-P2A-Puro	This study		ABE8e expressing lentiviral plasmid
Recombinant DNA reagent	pcDNA-3	Invitrogen		
Recombinant DNA reagent	pSG5/mKLF	***Miller and Bieker, 1993***		
Antibody	PE Mouse Anti-Human CD34 (mouse monoclonal)	BD Pharmingen	CAT # 550761; RRID:AB_393871	FACS (2 µl/test)
Antibody	APC Mouse Anti-Human CD133 (mouse monoclonal)	BD Pharmingen	CAT # 566596; RRID:AB_2744280	FACS (2 µl/test)
Antibody	BV421 Mouse Anti-Human CD90 (mouse monoclonal)	BD Pharmingen	CAT # 562556; RRID:AB_2737651	FACS (2 µl/test)
Antibody	PE-Cy7 Mouse Anti-Human CD235a (mouse monoclonal)	BD Pharmingen	CAT # 563666; RRID:AB_2738361	FACS (2 µl/test)
Antibody	BV421 Mouse Anti-Human CD71 (mouse monoclonal)	BD Pharmingen	CAT # 562995; RRID:AB_2737939	FACS (2 µl/test)
Antibody	PE Mouse Anti-Human CD235a (mouse monoclonal)	BD Pharmingen	CAT # 555570; RRID:AB_395949	FACS (2 µl/test)
Antibody	FITC Mouse Anti-Human CD71 (mouse monoclonal)	BD Pharmingen	CAT # 555536; RRID:AB_395920	FACS (2 µl/test)
Antibody	Fetal Hemoglobin Antibody, APC (mouse monoclonal)	Invitrogen	CAT # MHFH05; RRID:AB_10374595	FACS (2 µl/test)
Antibody	Antibody for KLF1 (rabbit polyclonal)	***Crossley et al., 1996***		EMSA (1:30 final dilution)

Appendix 1 Continued on next page

Appendix 1 Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	Anti-KLF1 antibody (goat polyclonal)	OriGene	#TA305808	ChiP (15 µg/IP)
Antibody	Normal rabbit IgG	Cell Signaling Technology	#2729S	ChiP (15 µg/IP)
Commercial assay or kit	NucRed Live 647 ReadyProbes Reagent	Invitrogen	CAT # R37106	FACS (2 µl/test)
Commercial assay or kit	Zymoclean Gel DNA recovery kit	Zymo Research	CAT # D4001	
Commercial assay or kit	NucleoBond Xtra Midi	MN	REF # 740410	
Commercial assay or kit	NucleoSpin RNA	MN	REF # 740955	
Commercial assay or kit	NucleoSpin Blood – DNA kit	MN	REF # 740951	
Commercial assay or kit	EasySep Human CD34 Positive Selection Kit	STEMCELL Technologies	CAT # 17856	
Commercial assay or kit	T7 mScript Standard mRNA Production System	Cell Script	C-MSC100625	
Commercial assay or kit	Radiolabelled DNA column	Roche	G25DNA-RO	
Commercial assay or kit	iScript cDNA Synthesis Kit	Bio-Rad	CAT # 1708891	
Commercial assay or kit	BigDye Terminator v3.1 Cycle Sequencing Kit	Applied Biosystem	CAT # 4337458	
Commercial assay or kit	SsoFast EvaGreen Supermix	Bio-Rad	CAT # 172–5200	
Commercial assay or kit	Universal Mycoplasma detection kit	ATCC	CAT # 30–1012K	
Chemical compound, drug	T4 Polynucleotide Kinase	NEB	CAT # M0201	
Chemical compound, drug	T4 DNA Ligase	NEB	CAT # M0202	
Chemical compound, drug	NEBuilder HiFi DNA Assembly Master Mix	NEB	CAT # E2621	
Chemical compound, drug	BamHI-HF	NEB	CAT # R3136	
Chemical compound, drug	NhEI-HF	NEB	CAT # R3131	
Chemical compound, drug	BsmB1	NEB	CAT # R0580	
Chemical compound, drug	KpnI-HF	NEB	CAT # R3142	
Chemical compound, drug	EcoRI-HF	NEB	CAT # R3101	
Chemical compound, drug	Exonuclease I (*E. coli*)	NEB	CAT # M0293	
Chemical compound, drug	Pme1	NEB	CAT # R0560	
Chemical compound, drug	GoTaq Green Master Mix	Promega	CAT # M712B	
Chemical compound, drug	PrimeSTAR GXL Premix	Takara Bio	CAT # R051A	
Chemical compound, drug	DynaBeads PG	Invitrogen	10003D	

Appendix 1 Continued on next page

Appendix 1 Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Chemical compound, drug	Formaldehyde	Sigma-Aldrich	F8775	1% v/v final concentration
Chemical compound, drug	Glycine	Ajax Finechem	AJA1083	125 mM final concentration
Chemical compound, drug	γ-[Kuzminov, 2001]P ATP	Perkin-Elmer	BLU502A250UC	1 µl/15 pmol probe
Chemical compound, drug	Insulin	Sigma-Aldrich	CAT # 11061-68-0	
Chemical compound, drug	Heparin	MP Biomedicals	CAT # 9041-08-1	
Chemical compound, drug	Holotransferrin	BBI Solutions	#SKU T101-5	
Chemical compound, drug	SCF	Immuno Tools	CAT # 11343325	
Chemical compound, drug	EPO	Zydus Nephrosciences	Zyrop 4000 IU Injection	
Chemical compound, drug	IL6	Immuno Tools	CAT # 11340066	
Chemical compound, drug	IL3	Immuno Tools	CAT # 11340035	
Chemical compound, drug	FLT3	Immuno Tools	CAT # 11343305	
Chemical compound, drug	TPO	Immuno Tools	CAT # 11344863	
Chemical compound, drug	Hydrocortisone	MP Biomedicals	CAT # 2930949	
Chemical compound, drug	AB Serum	MP Biomedicals	CAT # 101996	
Chemical compound, drug	Penstrep	Gibco	CAT # 15140122	
Chemical compound, drug	Dexamethasone	Alfa Aesar	CAS# 1177-87-3	
Chemical compound, drug	Doxycycline	Sigma-Aldrich	CAS# 24390-14-5	
Chemical compound, drug	Glutamine	Gibco	CAT # 25030081	
Chemical compound, drug	FBS	Gibco	CAT # 10270106	
Chemical compound, drug	PBS	Hyclone	CAT # SH30256.02	
Chemical compound, drug	Polybrene	Sigma-Aldrich	CAS # 28728-55-4	
Chemical compound, drug	Hepes	Gibco	CAT # 15630080	
Chemical compound, drug	Puromycin	Gibco	CAT # A1113803	
Chemical compound, drug	Pseudouridine	Jena Bioscience	CAT # NU-1139	
Chemical compound, drug	Lymphoprep	STEMCELL Technologies	CAT # 07851	
Chemical compound, drug	StemSpan SFEM-II	STEMCELL Technologies	CAT # 09655	
Chemical compound, drug	DMEM	Hyclone	CAT # SH30243.01	

Appendix 1 Continued on next page

Appendix 1 Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Chemical compound, drug	RPMI	Hyclone	CAT # SH30027.01	
Chemical compound, drug	IMDM-Glutamax	Gibco	CAT # 31980030	
Chemical compound, drug	Fugene HD	Promega Corporation	CAT # E2312	
Chemical compound, drug	Lenti-X concentrator	Takara	CAT # 631232	
Chemical compound, drug	Maxcyte buffer	Hyclone	CAT # EPB1	
Chemical compound, drug	LB Agar	HIMEDIA	M1151	
Chemical compound, drug	LB Broth	HIMEDIA	M1245	
Chemical compound, drug	Ampicillin Sodium Salt	SRL	61,314	
Chemical compound, drug	Triton X-100	Fisher Scientific	CAS #:9002931	
Chemical compound, drug	Glutaraldehyde	MP Biomedicals	CAT # 198,595	
Biological sample	PBMNCs	CMC	IRB Min. No. 10,549 (others) dated 15/02/2017	
Cell line (*Homo sapiens*)	HEK 293T	ATCC		
Cell line (*Homo sapiens*)	HUDEP-2	Cell Engineering Division, RIKEN BioResource Center		
Cell line (*Homo sapiens*)	K562	ATCC		
Cell line (African green monkey)	COS-7	***Gluzman, 1981***		
Sequence-based reagent	gRNAs	This paper		Check ***Supplementary file 1***
Sequence-based reagent	Probes, RT-qPCR and PCR primers	This paper		Check ***Supplementary file 2***
Software, algorithm	Reditools 2	GitHub - BioinfoUNIBA/REDItools2, ***Giudice, 2022***		RNA off-target
Software, algorithm	Synthego ICE	Synthego		InDel for Sanger sequenced data
Software, algorithm	EditR	EditR: Edit Deconvolution by Inference of Traces in R (shinyapps.io)		Base editing efficiency for Sanger sequenced data
Software, algorithm	IGV	Home | Integrative Genomics Viewer (broadinstitute.org)		Visualize Aligned data
Software, algorithm	CRISPResso-2	CRISPResso2 (partners.org)		Base editing efficiency for NGS data
Software, algorithm	Snapgene	SnapGene | Software for everyday molecular biology		gRNA designing
Software, algorithm	Benchling	CRISPR gRNA Design Tool | Benchling		gRNA designing
Software, algorithm	FlowJo 10.7.1	Home | FlowJo, LLC		FACS data analysis

Appendix 1 Continued on next page

Appendix 1 Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	Cas off-finder	CRISPR RGEN Tools (rgenome.net)		DNA off-target prediction
Software, algorithm	Cosmid	CRISPR Target Search (gatech.edu)		Primer designing for predicted DNA off-targets
Software, algorithm	Bowtie 2	Bowtie 2: fast and sensitive read alignment (sourceforge.net)		Sequence alignment
Software, algorithm	TrimGalore	Babraham Bioinformatics - Trim Galore!		FastQ files processing
Software, algorithm	NFCore RNA Seq pipeline	rnaseq » nf-core		RNA sequencing pipeline
Software, algorithm	Interactive Gene Expression Analysis Kit (iGEAK)	iGEAK! (google.com)		
Software, algorithm	GraphPad Prism 8.0.1	GraphPad Prism (https:// graphpad.com)	RRID:SCR_015807	

Appendix c - Plagiarism check report

Institute for Stem Cell Science and Regenerative Medicine

An autonomous Institutee of
Department of Biotechnology,
Government of India.

Center for Stem Cell Research

(A unit of inStem, Bengaluru in Collaboration with DBT and CMC, Vellore)

CERTIFICATE

This is to certify that the thesis submitted by **Mr. NITHIN SAM. R**, affiliated with the Center for Stem Cell Research at Christian Medical College, has undergone a thorough plagiarism check utilizing iThenticate software. We are pleased to declare that the thesis is deemed acceptable for submission. While the overall similarity index was 20%, it is noteworthy that the individual primary source index fell within the permissible range of 1-5%.

Title: **GENOME EDITING STRATEGIES FOR THE TREATMENT OF HEREDITARY HEMATOLOGICAL DISORDERS**

Registration number: **2019/PhD/08**

Name of the Guide: **Dr. Mohankumar K. Murugesan**

Date: 10-06-2023

Dr. Shaji R.V

CENTRE FOR STEM CELL RESEARCH
Christian Medical College Campus, Bagayam, Vellore - 632 002, Tamil Nadu, India.
A unit of inStem, Bengaluru

Christian Medical College Campus,
Bagayam, Vellore - 632002. INDIA.

Phone : +91 - 416 - 228 5101 / 5107
Fax : +91 - 416 - 228 5103
E-mail : cscr@cmcvellore.ac.in