Université Paris Cité

École doctorale Hématologie, Oncogenèse et Biothérapies : ED 561 HOB

IHU Imagine - Institut des maladies génétiques, Inserm U1163
Équipe Lymphohématopoïèse humaine

Base Editing Approaches for the Treatment of β-hemoglobinopathies

Par Giulia HARDOUIN

Thèse de doctorat de Biothérapies et biotechnologies

Dirigée par Pr. Marina CAVAZZANA

Présentée et soutenue publiquement à Paris le 5 Décembre 2022

Devant un jury composé de :

Pr. John STROUBOULIS	Full professor	King's College London	Rapporteur
Dr. Paula RIO	Group leader	CIEMAT, Madrid	Rapportrice
Dr. Annarita MICCIO	CR-HDR	Université Paris Cité	Examinatrice
Dr. Carine GIOVANNANGELI	DR	MNHN, Paris	Examinatrice
Dr. Mario AMENDOLA	CR-HDR	Université Paris Saclay	Examinateur
Pr. Marina CAVAZZANA	PU-PH	Université Paris Cité	Directrice de thèse

Approches d'édition de base pour le traitement des β-hémoglobinopathies.

Les β-hémoglobinopathies sont dues à des mutations affectant la synthèse ou la structure des chaînes de la β-globine constituant l'hémoglobine adulte (HbA). La β-thalassémie est une maladie récessive très répandue causée par des mutations dans le locus du gène de la β-globine (*HBB*) qui réduisent (β^+) ou éliminent (β^0) la production de β-globine. Les chaînes d'α-globine découplées précipitent et provoquent l'apoptose des précurseurs érythroïdes et une érythropoïèse inefficace. Selon le génotype, le phénotype clinique peut varier d'une anémie légère à une anémie dépendante des transfusions. La drépanocytose est causée par une mutation ponctuelle unique dans le gène *HBB*, qui entraîne la production d'une globine β^S défectueuse. Les tétramères d'Hb drépanocytaire (HbS) qui en résultent polymérisent sous hypoxie, déforment les globules rouges et provoquent une obstruction des micro-vaisseaux conduisant à une maladie chronique multi-organes. La transplantation de cellules souches/progénitrices hématopoïétiques (HSPCs) autologues génétiquement modifiées est une option thérapeutique intéressante. Cependant, les stratégies actuelles de thérapie génique basées sur l'utilisation de vecteurs lentiviraux ou de la nucléase CRISPR/Cas9 ne sont pas également efficaces chez tous les patients et/ou posent des problèmes de sureté. Les éditeurs de bases adénines (ABE), un outil d'édition du génome dérivé de CRISPR/Cas9, permettent l'insertion de mutations ponctuelles A>G à une position spécifique du génome.

Les deux mutations β^0 CD39 (CAG>TAG) et IVS2-1 (G>A) et la mutation β^+ IVS1-110 (G>A) sont parmi les mutations β-thalassémiques les plus courantes et les plus sévères de la région méditerranéenne et du Moyen-Orient. Nous avons exploité la capacité des ABEs en combinaison avec des ARN guides (gRNA) spécifiques pour corriger ces mutations. Pour chaque mutation, nos combinaisons ABE/gRNA ont permis d'obtenir des efficacités de correction allant jusqu'à ~90 % dans les HSPCs de patients atteints de β-thalassémie. Les HSPC β-thalassémiques modifiées ont été différenciées vers la lignée érythroïde pour évaluer la production de globine et d'Hb. Les globules rouges dérivés des HSPC modifiées ont présenté des niveaux élevés de β-globine et une amélioration des ratios α/non-α globine. Le retard de différenciation érythroïde ainsi que le profil apoptotique et les niveaux élevés de stress oxydatif typiquement observés dans les cultures de cellules β-thalassémiques ont été corrigés par notre traitement. En effet, au cours de la différenciation, nous avons observé un taux d'énucléation accru par rapport aux contrôles β-thalassémiques non édités, atteignant les niveaux observés dans des cellules de donneurs sains.

Avec les éditeurs de bases actuels, la mutation T>A responsable de la drépanocytose ne peut pas être inversée. Cependant, les ABEs peuvent convertir le A mutant en G (A>G). L'insertion de cette mutation crée un variant allélique connu de *HBB*, non pathogène, conduisant à la production de l'HbG-Makassar. Nous avons utilisé une lignée cellulaire portant la mutation SCD, SCD-HUDEP2, afin de sélectionner les combinaisons ABEs/gRNAs les plus performantes. Comme la mutation drépanocytaire était inaccessible avec la première génération d'ABEs, nous avons successivement modifié les domaines de l'enzyme (désaminase, linker et Cas9) et optimisé la longueur des gRNAs afin d'accéder au locus et d'augmenter l'efficacité de l'édition de bases sur la mutation. Ainsi, nous avons réussi à corriger

jusqu'à ~80% des allèles dans des HSPCs de patient drépanocytaire. La suite du projet visera à démontrer le potentiel thérapeutique de cette stratégie dans les globules rouges dérivés des HSPCs éditées et à étudier les propriétés biophysiques de l'HbG-Makassar.

En conclusion, nous avons développé des approches efficaces d'édition de bases pour corriger trois mutations β-thalassémiques sévères et hautement prévalentes ainsi que la mutation drépanocytaire. Les études de génotoxicité et la validation de ces résultats dans des HSPCs *in vivo* fourniront une preuve suffisante de sureté et d'efficacité pour permettre le développement clinique des stratégies par édition de bases des HSC pour la thérapie génique des β-hémoglobinopathies.

Mots clés : β-hémoglobinopathies, drépanocytose, β-thalassémie, thérapie génique, édition de base

Base editing approaches for the treatment of β-hemoglobinopathies.

β-hemoglobinopathies are caused by mutations affecting the synthesis or the structure of the β-globin chain of the adult hemoglobin (HbA). β-thalassemia is a highly prevalent recessive disorder caused by mutations in the β-globin gene (*HBB*) locus that reduce (β^+) or eliminate (β^0) β-globin production. The uncoupled α-globin chains precipitate and cause apoptosis of erythroid precursors and inefficient erythropoiesis. Depending on the genotype, the clinical phenotype can vary from mild to transfusion-dependent anemia. Sickle cell disease (SCD) is caused by a single point mutation in *HBB*, which leads to the production of the defective β^S-globin. The resulting Hb tetramers (HbS) polymerize under hypoxia, deform red blood cells (RBCs) and cause micro vessels obstruction leading to a chronic multi-organ disease. Transplantation of autologous, genetically modified hematopoietic stem/progenitor cells (HSPCs) is an attractive therapeutic option. However, current gene therapy strategies based on the use of lentiviral vectors or CRISPR/Cas9 nuclease are not equally effective in all the patients and/or raise safety concerns. Adenine base editors (ABEs), a CRISPR/Cas9-based genome editing tool, allow the insertion of A>G point mutations at a specific position.

The two β^0 mutations CD39 (CAG>TAG) and IVS2-1 (G>A) and the β^+ mutation IVS1-110 (G>A) are among the most common and severe β-thalassemia mutations in the Mediterranean area and Middle East. Here, we exploited the capacity of ABEs in combination with specific single guide RNAs (gRNAs) to correct these mutations. For each mutation, our ABEs/gRNAs combinations achieved correction efficiencies of up to ~90% in HSPCs from β-thalassemia patients. Edited β-thalassemic HSPCs were differentiated towards the erythroid lineage to evaluate globin and Hb production. RBCs derived from edited HSPCs exhibited high β-globin levels and improvement of the α/non-α globin ratios. The delayed erythroid differentiation as well as the apoptotic profile and elevated ROS levels typically observed in β-thalassemic cell cultures were corrected by our treatment. Indeed, along the differentiation, we observed an increased enucleation rate compared to unedited β-thalassemic controls, reaching the levels observed in healthy donor cells.

With current base editors, the SCD-causing T>A mutation cannot be reverted. However, ABEs can convert the mutant A into a G (A>G). Insertion of this mutation creates a known non-pathogenic *HBB* allelic variant, leading to HbG-Makassar production. We used SCD-HUDEP2, a cell line carrying the SCD mutation, in order to select the best-performing ABEs/gRNAs combinations. As the SCD mutation was inaccessible with the first generation of base editors, we engineered ABE domains (deaminase, linker and Cas9) and optimized the gRNAs length in order to access the locus and increase base editing efficiencies at the target base. Finally, we achieved up to ~80% correction of the SCD mutation in SCD HSPCs. Future experiments aim to demonstrate the therapeutic potential of this strategy in RBCs derived from edited cells and to study biophysical properties of HbG-Makassar.

In conclusion, we developed efficient base editing approaches to revert three highly prevalent and severe β-thalassemia mutations and the SCD mutation. Genotoxicity studies and validation of these results in HSPCs *in vivo* will provide sufficient proof of safety and efficacy to enable the clinical development of base-edited HSCs for gene therapy of β-hemoglobinopathies.

Keywords : β-hemoglobinopathies, sickle cell disease, β-thalassemia, gene therapy, base editing

Abbreviations

A	adenine
AAG	alkyl adenine DNA glycosylase
ABE	adenine base editor
BER	base excision repair
BFU-E	burst-forming unit-erythroid
C	cytosine
Cas	CRISPR-associated
CBE	cytosine base editor
CFU-E	colony-forming unit-erythroid
CLP	common lymphoid progenitor
CMP	common myeloid progenitor
CRISPR	clustered regularly interspaced short palindromic repeats
dCas9	dead Cas9
DDR	DNA damage response
DSB	double strand break
EB	erythroblast
EMA	European Medicines Agency
FDA	Food and Drug Administration
G	guanine
GMP	granulocyte-monocyte progenitor
gRNA	single guide RNA
Hb	hemoglobin
HbA	adult hemoglobin
HbA2	adult hemoglobin 2
HbF	fetal hemoglobin
HbG	hemoglobin Makassar
HDR	homologous directed repair
HPFH	hereditary persistence of fetal hemoglobin
HSC	hematopoietic stem cell
HSPC	hematopoietic stem/progenitor cell
I	inosine
InDel	insertion/deletion
kb	kilobase
LCR	locus control region
LMPP	lympho-myeloid primed progenitor
LT-HSC	long-term hematopoietic stem cell
MEP	megakaryocyte-erythroid progenitor
MFI	mean fluorescence intensity
MMEJ	microhomology-mediated end joining
MMR	DNA mismatch repair
MPP	multipotent progenitor
mRNA	messenger RNA
nCas9	Cas9 nickase
NGS	next generation sequencing
NHEJ	non-homologous end-joining
NO	nitric oxide
PAM	protospacer adjacent motif
pre-crRNA	pre-CRISPR RNA

RBC	red blood cell
RNP	ribonucleoprotein
ROS	reactive oxygen species
SEM	standard error of the mean
SCD	sickle cell disease
SSB	single strand break
ssDBD	single-strand DNA-binding domain
Sp	*Streptococcus pyogenes*
T	thymine
TALEN	transcription-activator-like effector nucleases
tracrRNA	trans-activator RNA
TSS	transcription start site
U	uracil
UGI	uracil glycosylase inhibitor
UNG2	nuclear uracil DNA N-glycosylase
VCAM-1	vascular cell adhesion molecule-1
WES	whole exome sequencing
WT	wild-type
ZFN	zinc-finger nucleases

Table of Contents

ABBREVIATIONS **6**

TABLE OF CONTENTS **8**

RÉSUMÉ **10**

I. INTRODUCTION **16**

A. HEMOGLOBIN AND GLOBIN GENE REGULATION 16
1. Role and structure *16*
2. Globin gene regulation *17*
3. Hereditary Persistence of Fetal Hemoglobin *18*
B. RED BLOOD CELLS PRODUCTION 21
1. Hematopoiesis *21*
2. Erythropoiesis *22*
C. SICKLE CELL DISEASE 23
1. Epidemiology *23*
2. Pathophysiology *23*
3. Clinical manifestations *25*
4. Treatments *26*
D. Β-THALASSEMIA 28
1. Epidemiology *28*
2. Pathophysiology *28*
3. Clinical manifestations *31*
4. Treatments *32*
E. CRISPR/CAS9 32
1. CRISPR/Cas9 in prokaryotic organisms *33*
2. CRISPR/Cas9 as a genome editing tool *35*
3. Safety of CRISPR/Cas9 usage *37*
a) Editing byproducts at the on-target site 37
b) Editing byproducts at the off-target sites 38
c) DNA Damage Response 38
F. BASE EDITING 38
1. Development of base editors *38*
2. Genome editing with base editors: requirements and optimizations *41*
3. Safety of the base editing procedure *43*
a) Editing byproducts at the on-target site 43
b) Editing byproducts at the off-target sites 45
G. GENE THERAPY STRATEGIES FOR THE TREATMENTS OF Β-HEMOGLOBINOPATHIES 45
1. Gene therapy: rationale and principle *45*
2. Lentivirus-based strategies *48*
3. Genome editing strategies *49*
a) Strategies aiming to correct the SCD mutation 50
(1) CRISPR/Cas9-based approaches 50
(2) Base editing-based approaches 51
b) Strategies aiming to reactivate HbF expression 52
(1) Targeting the *BCL11A* erythroid enhancer 53
(a) CRISPR/Cas9-based approaches 53
(b) Base editing-based approaches 54
(2) Targeting regulators binding sites in the *HBG* genes 55
(a) CRISPR/Cas9-based approaches 55
(b) Base editing-based approaches 56

II. AIM OF THE PROJECT **58**

III. ADENINE BASE EDITOR-MEDIATED CORRECTION OF A COMMON AND SEVERE Β-THALASSEMIA MUTATION **60**

A. ABSTRACT 61
B. INTRODUCTION 63

C. Materials and Methods 64
D. Results and discussion 64
E. Figures 67
F. Figure legends 70
G. Supplemental files 73
1. Supplemental figures 73
2. Supplemental figure legends 79
3. Supplemental discussion 80
4. Supplemental materials and methods 81
5. Supplemental Tables 85

IV. BASE EDITING APPROACHES FOR THE TREATMENT OF TWO B^0 THALASSEMIA MUTATIONS. 87

A. Introduction 88
B. Results 91
C. Discussion 94
D. Figures 96
E. Figure legends 100
F. Supplemental files 104
1. Supplemental figures 104
2. Supplemental figure legends 106
G. Materials and Methods 107

V. DEVELOPMENT OF A BASE EDITING STRATEGY TO CORRECT THE SCD MUTATION. 113

A. Introduction 114
B. Results 115
C. Discussion 119
D. Figures 122
E. Figure legends 126
F. Materials and Methods 128

VI. DISCUSSION AND PERSPECTIVES 132

A. Genome editing approaches for β-hemoglobinopathies 132
B. Editing in long-term repopulating HSCs 134
C. Off-target editing 135
D. Bystander editing 136
E. Prime editing 136
F. Expected phenotypic correction using these approaches 137
G. The challenge of costs and accessibility to gene therapy 138
H. Conclusions 140

ACKNOWLEDGMENTS 141

BIBLIOGRAPHY 144

ANNEX 1: BASE-EDITING-MEDIATED DISSECTION OF A γ-GLOBIN CIS-REGULATORY ELEMENT FOR THE THERAPEUTIC REACTIVATION OF FETAL HEMOGLOBIN EXPRESSION 157

Résumé

Les β-hémoglobinopathies sont dues à des mutations affectant la synthèse ou la structure des chaînes de la β-globine constituant l'hémoglobine adulte (HbA). La β-thalassémie est une maladie récessive très répandue affectant environ 60 000 enfants chaque année, principalement en Asie du sud-est, Méditerranée, Moyen Orient et Afrique subsaharienne. Cette maladie est causée par des mutations dans le locus du gène de la β-globine (HBB) qui réduisent (β+) ou éliminent (β0) la production de β-globine. Les chaînes d'α-globine découplées précipitent et provoquent l'apoptose des précurseurs érythroïdes et une érythropoïèse inefficace. L'hétérogénéité des phénotypes reflète l'hétérogénéité moléculaire et allélique de ce trouble, qui entraîne différents degrés de déséquilibre dans la production des chaînes α et β. La β-thalassémie majeure, qui inclut les sujets homozygotes pour les mutations β0 ou hétérozygotes pour les mutations β0 et β+ sévères, se présente sous la forme d'une profonde anémie hémolytique microcytaire avec une dépendance aux transfusions.

La drépanocytose est causée par une mutation ponctuelle unique dans le gène HBB, qui entraîne la production d'une globine βS défectueuse. Les tétramères d'Hb drépanocytaire (HbS) qui en résultent polymérisent sous hypoxie, déforment les globules rouges et provoquent une obstruction des micro-vaisseaux conduisant à une maladie chronique multi-organes. D'autre part, les cycles de falciformation-défalciformation affaiblissent les globules rouges et entraînent une hémolyse. La drépanocytose se caractérise par la combinaison d'une anémie hémolytique chronique et de complications aiguës telles que les crises vaso-occlusives. Les crises vaso-occlusives se manifestent par des douleurs très intenses d'apparition brutale, pouvant être responsables de la nécrose des organes touchés et l'altération de leur fonction à long terme (déficits neurologiques, rétinopathie, cardiomyopathie, ostéonécrose, insuffisance rénale chronique, impuissance...). Ainsi, les manifestations chroniques sont dues à l'association de l'anémie hémolytique et des altérations fonctionnelles des organes touchés par les crises vaso-occlusives.

La transplantation de cellules souches/progénitrices hématopoïétiques (HSPCs) autologues génétiquement modifiées est une option thérapeutique intéressante.

Cependant, malgré des résultats très encourageants, les stratégies actuelles d'addition de gène basées sur l'utilisation de vecteurs lentiviraux se heurtent à certaines limites. En effet, plus le phénotype est sévère, plus la réponse est variable. Par exemple, il semblerait que l'Hb résultant de l'expression du transgène soit insuffisante chez les patients atteints de β-thalassémie β0/β0 et qu'ils nécessiteraient la transduction d'un plus grand nombre de copies de vecteurs par cellule pour obtenir les mêmes résultats que ceux obtenus chez les patients non-β0/β0. Ainsi, les efforts se concentrent sur l'optimisation des efficacités de transduction pour garantir un grand nombre de cellules modifiées et des niveaux élevés d'expression du transgène afin d'obtenir des résultats plus homogènes. Cependant, l'augmentation des événements d'intégration du transgène augmente théoriquement le risque de génotoxicité. En effet, l'intégration d'une séquence exogène dans le génome de l'hôte peut déclencher une oncogenèse en activant des proto-oncogènes. Cet événement indésirable est appelé mutagenèse insertionnelle.

Aujourd'hui, de nouvelles stratégies d'édition du génome, telles que CRISPR/Cas9, offrent la possibilité de corriger les mutations responsables des β-hémoglobinopathies et de restaurer l'expression endogène de l'HBB, ou d'induire des niveaux élevés d'expression endogène d'hémoglobine fœtale (HbF). Alors que les lentivirus ne peuvent pas inclure la totalité de la séquence régulatrice de l'HBB en raison de la limite d'encapsulation, la réexpression des gènes de globines endogènes médiée par l'édition du génome a l'avantage de bénéficier de l'environnement génétique physiologique du gène et donc de permettre un niveau plus élevé d'expression thérapeutique de la globine. De plus, ces stratégies ne nécessitent pas l'insertion d'ADN exogène, évitant ainsi le risque de mutagenèse insertionnelle.

Cependant, ces stratégies, basées sur l'utilisation de nucléases, génèrent des coupures doubles brins génotoxiques. Ainsi, la correction d'une séquence s'accompagne de la génération d'insertions et délétions (InDels) au niveau du site de coupure, présentant une limite à l'efficacité de la réparation. Plus récemment, les mêmes stratégies ont été explorées en utilisant des outils d'édition du génome sans coupure double brin, tels que les éditeurs de bases, qui peuvent être très efficaces dans les cellules souches hématopoïétiques (CSHs). C'est dans ce contexte que

nous avons cherché à développer de nouvelles stratégies de thérapie génique plus sures pour les β-hémoglobinopathies.

Les éditeurs de bases adénines (ABE), un outil d'édition du génome dérivé de CRISPR/Cas9, permettent l'insertion de mutations ponctuelles A>G à une position spécifique du génome. Les éditeurs de bases ont été développés en combinant une nucléase Cas9 inactivée (nCas9) avec une désaminase. Cet outil présente l'avantage de ne pas générer de coupure double brin, et ne nécessite pas l'administration d'une matrice d'ADN exogène, ce qui en fait une stratégie plus sûre et moins complexe pour corriger les mutations.

La β-thalassémie peut résulter d'une grande variété de mutations incluant des substitutions, de petites insertions et des délétions. Parmi les mutations les plus répandues, nous avons identifié trois mutations sévères pouvant être corrigées par édition de bases. Les deux mutations β0 CD39 (CAG>TAG) et IVS2-1 (G>A) et la mutation β+ IVS1-110 (G>A) sont parmi les mutations β-thalassémiques les plus courantes et les plus sévères de la région méditerranéenne et du Moyen-Orient. Nous avons exploité la capacité des ABEs en combinaison avec des ARN guides (gRNA) spécifiques pour corriger ces mutations dans les HSPCs de patients β-thalassémiques.

Dans un premier temps, nous avons criblé différentes combinaisons de gRNA/ABE dans des lymphocytes T obtenus par prélèvement de sang périphérique de patients β-thalassémiques présentant les mutations ciblées. Les cellules ont été transfectées avec des gRNAs chimiquement modifiés et des ARNm d'ABE transcrits in vitro. Ce premier criblage a permis d'identifier la combinaison de gRNA/ABE la plus efficace pour corriger les différentes mutations β-thalassémiques ciblées. Dans tous les cas, nous avons observé une génération d'InDels faible ou négligeable.

Dans un second temps, nous avons transfectés des HSPCs - la population cible dans le cas d'une thérapie génique - isolées à partir de prélèvements de sang périphérique de patients β-thalassémiques présentant les mutations ciblées. Pour chaque mutation, nous avons transfecté ces cellules avec la combinaison de gRNA/ABE sélectionnée lors de l'étape de criblage dans les lymphocytes T. Nous

avons utilisé un protocole d'édition cliniquement pertinent consistant en la transfection de l'ARNm de l'ABE transcrit in vitro, en parallèle avec un gRNA chimiquement modifié.

Pour chaque mutation, nos combinaisons ABE/gRNA ont permis d'obtenir des efficacités de correction allant jusqu'à ~90 % dans les HSPCs de patients atteints de β-thalassémie. Les HSPCs β-thalassémiques modifiées ont été différenciées vers la lignée érythroïde pour évaluer la production de globine et d'Hb. Alors que le mécanisme de dégradation des ARNm non-sens est connu pour dégrader rapidement les ARNm de β-globine porteurs de la mutation β0 CD39 (CAG>TAG), nos résultats montrent la restauration de niveaux normaux d'ARNm fonctionnels. De plus, la correction des mutations IVS2-1 (G>A) et IVS1-110 (G>A) conduit à la correction respective des défauts d'épissage et à la production de niveaux normaux d'ARNm de β-globine fonctionnels correctement épissés. Ainsi, les globules rouges dérivés des HSPCs modifiées ont présenté des niveaux élevés de β-globine et une amélioration des ratios α/non-α globine. La correction du phénotype a été évaluée par cytométrie de flux. Le retard de différenciation érythroïde ainsi que le profil apoptotique et les niveaux élevés de stress oxydatif typiquement observés dans les cultures de cellules β-thalassémiques ont été corrigés par notre traitement. En effet, au cours de la différenciation, nous avons observé un taux d'énucléation accru par rapport aux contrôles β-thalassémiques non édités, atteignant les niveaux observés dans des cellules de donneurs sains. De plus, à la fin de la différenciation, la taille des cellules énucléées (typiquement réduite dans les cellules β-thalassémiques en culture) a augmenté dans les échantillons édités.

Par ailleurs, nous avons évalué le profil de sécurité de la stratégie visant à corriger la mutation IVS1-110 (G>A), par description d'une activité hors cible limitée à une région dépourvue de gènes connus et sans impact sur la viabilité et la différenciation des cellules éditées. De plus, la correction de la mutation entrainant l'édition de bases situées à proximité de la base ciblée (située dans l'intron 1 de HBB), nous avons confirmé l'absence d'impact de ces mutations sur l'expression du gène HBB. Enfin, nous avons démontré l'absence de toxicité du traitement sur la viabilité des progéniteurs érythroïdes et granulo-monocytaires.

Enfin, nous avons utilisé la microscopie de phase pour évaluer de manière unicellulaire et quantitative la correction de plusieurs paramètres physiques des globules rouges suite à la correction de la mutation IVS1-110 (G>A). Nous avons pu décrire que les globules rouges obtenus après différentiation érythroïde in vitro des HSPCs β-thalassémiques étaient caractérisés par un périmètre, une surface et un volume réduits en raison de la microcytose, et une masse sèche et une densité de surface diminuées en raison de la faible teneur en Hb par rapport à des globules rouges sains. Ces paramètres ont été normalisés ou considérablement améliorés dans les conditions β-thalassémiques traitées.

Avec les éditeurs de bases actuels, la mutation T>A responsable de la drépanocytose ne peut pas être corrigée. Cependant, les ABEs peuvent convertir le A mutant en G (A>G). L'insertion de cette mutation crée un variant allélique connu de HBB, non pathogène, conduisant à la production de l'HbG-Makassar. Nous avons utilisé une lignée cellulaire portant la mutation SCD, SCD-HUDEP2, afin de sélectionner les combinaisons ABEs/gRNAs les plus performantes. Comme la mutation drépanocytaire était inaccessible avec la première génération d'ABEs, nous avons successivement modifié les domaines de l'enzyme (désaminase, connecteur et Cas9) et optimisé la longueur des gRNAs afin d'accéder au locus et d'augmenter l'efficacité de l'édition de bases sur la mutation. Alors qu'aucun des paramètres optimisés ne permettait à lui seul de dépasser 30% de correction du gène, la combinaison des différents paramètres optimisés a permis de corriger jusqu'à ~80% des allèles dans des HSPCs de patient drépanocytaire. Ainsi, en plus du développement d'une stratégie d'édition de base efficace pour la conversion de la mutation drépanocytaire, ce travail rassemble certaines des optimisations les plus récentes des outils d'édition des bases et démontre comment elles peuvent être combinées afin d'atteindre efficacement les positions difficiles à éditer. La suite du projet visera à démontrer le potentiel thérapeutique de cette stratégie dans les globules rouges dérivés des HSPCs éditées et à étudier les propriétés biophysiques de l'HbG-Makassar.

En conclusion, nous avons développé des approches efficaces d'édition de bases pour corriger trois mutations β-thalassémiques sévères et hautement prévalentes ainsi que la mutation drépanocytaire. Les études de génotoxicité et la validation de

ces résultats dans des HSPCs in vivo fourniront une preuve suffisante de sureté et d'efficacité pour permettre le développement clinique des stratégies par édition de bases des HSCs pour la thérapie génique des β-hémoglobinopathies.

I. Introduction

A. Hemoglobin and globin gene regulation

1. Role and structure

Red blood cells (RBCs) are responsible for oxygen transportation throughout the body. Their main component, hemoglobin (Hb), is a tetrameric protein composed of two α- and two β-like globins (**Figure 1**). Each globin is associated with a heme molecule, which contains an iron that can bind or release an oxygen molecule (**Figure 1**). A precise physicochemical regulation allows oxygen capture in the lungs, followed by transportation and release in the tissues. On the contrary, byproducts of cellular respiration, CO_2 and H^+, are carried back to the lungs.

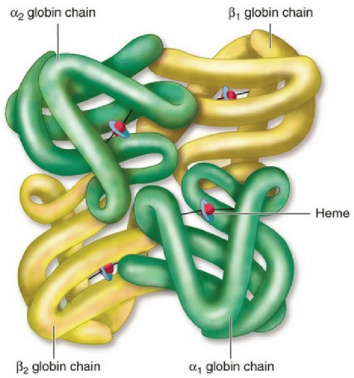


Figure 1. Hemoglobin: a hemoprotein responsible for gas exchange in red blood cells.

Hemoglobin is composed of two α- and two β-like chains, each associated with a heme molecule (*Adapted from Mettananda et al. 2015*[1]).

In humans, β-like globins are sequentially expressed during life. The first ones to be expressed are ε- and γ-chains during the embryonic and the fetal stage respectively, and soon after birth they are replaced by β- and δ-chains (**Figure 2**). This phenomenon is called Hb switching. The major Hb in the adult life is the adult Hb (HbA), which consists of 2 α- and 2 β-chains (HbA: α2β2) while the adult Hb 2 (HbA2: α2δ2) and fetal Hb (HbF: α2γ2) are expressed at low levels (**Figure 2**). Similarly, the α-like ζ-globin is expressed during embryonic life and is replaced by α-globin after the first 3 months of gestation.

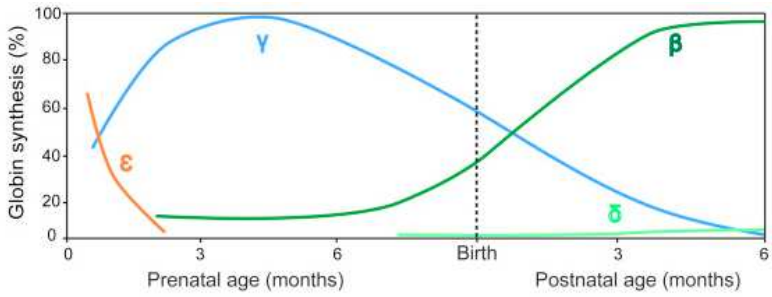


Figure 2. Expression of β-like globin chains during human development.

β-like globin chains expression is sequential. While ε- and γ-chains are the main β-like globins expressed before birth, they are replaced soon after birth by β- and δ-globins. (*Adapted from Cavazzana et al., 2017*[2])

2. Globin gene regulation

Globin genes are clustered within two loci. β-like globin genes are located on chromosome 11, whereas α-like globin gene locus is located on chromosome 16. The latter is constituted by *HBZ*, *HBA2,* and *HBA1*, encoding for ζ-, α2-, and α1-globins, respectively. Both α2- and α1- globins are expressed at the adult stage and encode for the same α-globin protein[3]. Regarding the β-locus, it contains *HBE1*, *HBG2*, *HBG1*, *HBD*, and *HBB*, encoding for ε-, γG-, γA-, δ-, and β-globin respectively. Of note, the two γ-globin genes (*HBG1* and *HBG2*) are expressed and encode for two γ-globin chains differing by one amino acid in position 136. β-like genes sequential expression is regulated by the locus control region (LCR). This 16-kb long region contains 5 DNaseI hypersensitive sites, i.e., active regulatory regions (**Figure 3**). The LCR allows chromatin opening and gene expression by direct interaction with the β-like gene promoters. β-like globin activation is mediated by a dynamic loop that evolves during development to make the LCR interact with *HBG* promoters during fetal life and with the *HBB/HBD* genes after birth (**Figure 3**). This long-range interaction is thought to be mediated by a pentameric complex composed of the proteins GATA1-TAL1-E2A-LMO2-LDB1 and stabilized by the self-interacting domain of LDB1[2,4] (**Figure 3**).

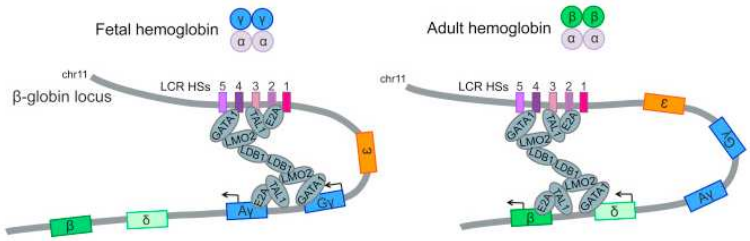


Figure 3. Locus control region and regulation of β-like globin genes.

On the left, before birth, the LCR interacts with *HBG1/2* promoters. On the right, in adulthood, the LCR interacts with *HBB/HBD* promoters. (*Adapted from Cavazzana et al., 2017*[2].)

3. Hereditary Persistence of Fetal Hemoglobin

Hereditary Persistence of Fetal Hemoglobin (HPFH) is a condition in which elevated HbF levels persist during adulthood. While being benign in healthy individuals, this condition is of particular interest when found in individuals with sickle cell disease (SCD) or β-thalassemia as elevated HbF levels significantly reduces the morbidity and mortality of these diseases[5,6]. This protective effect might be thanks to the anti-sickling properties of HbF in SCD or its ability to compensate for the lack of functional Hb in β-thalassemia[2].

The study of the molecular basis of this condition has shown that there are two HPFH categories according to the type of causative mutation:
(i) Deletional HPFH forms are caused by large deletions in the β-locus. Increased expression of γ-chains potentially results from the juxtaposition of enhancer-type regions with *HBG1/2* promoters or the loss of inhibitor sequences[7,8].
(ii) Non-deletional HPFH forms result from point mutations which generally affect the *HBG1/2* promoters. These mutations alter the binding sites of transcriptional repressors or, conversely, create binding sites for transcription factors that positively regulate γ-globin expression. Systematic studies of HPFH point mutations impact on HbF expression helped to elucidate some key mechanisms of HbF regulation.

HPFH point mutations are clustered in 3 regions of the homologous *HBG1* and *HBG2* promoters. These regions are located -200, -175, and -115 base pairs upstream of the transcription start site (TSS) (**Figure 4**).

The -200 and -115 regions contain binding sites for two potent HbF repressors, LRF and BCL11A, respectively. HPFH mutations altering these binding sites are responsible for the eviction of these repressors thus allowing the maintenance of HbF expression[9,10] (**Figure 4**).

Other HPFH mutations cause the recruitment of transcription activators to the *HBG1/2* promoters. For instance, the -198 (T>C), -175 (T>C) and -113 (A>G) mutations create binding sites for the transcription factors KLF1, TAL1 and GATA1, respectively[11–13] (**Figure 4**).

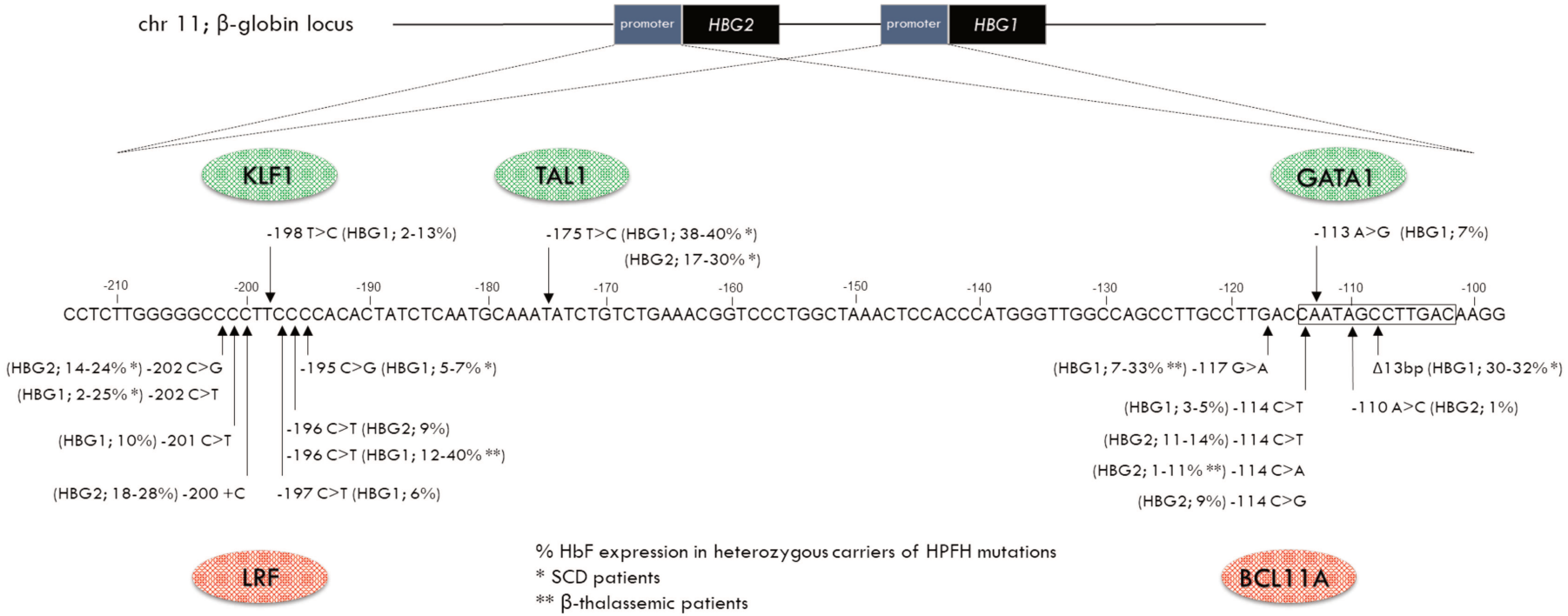


Figure 4. HPFH point mutations in the *HBG1/2* promoters.

HPFH can be due to the presence of point mutations in the promoters of the γ-globin genes *HBG1* and *HBG2*. These mutations alter repressor binding sites (e.g., LRF or BCL11A) or conversely create binding sites for transcription factors that positively regulate γ-globin expression (e.g., KLF1, TAL1, or GATA1). (*Adapted from Martyn et al., 2017*[14])

B. Red blood cells production

1. Hematopoiesis

All blood cells are obtained by the differentiation of hematopoietic stem cells (HSCs) (**Figure 5A**). HSCs are multipotent, as they can differentiate and produce the 3 hematopoietic lineages: erythrocytes, leukocytes, and thrombocytes. Moreover, they are able to self-renew, which ensures the sustainability of this production throughout life. Hematopoiesis is often represented as a succession of phenotypically distinct cell states, each of them corresponding to progressive lineage commitment steps (**Figure 5A**). Interestingly, recent insights from novel lineage tracing and single-cell transcriptomic studies suggest that cell state transitions might occur as a continuum of transitory states rather than as a succession of irreversible discrete states[15] (**Figure 5B**). Moreover, it appears that progenitor commitment might be relative. Thus, even though some cell state transitions are more likely to occur than others, already lineage-restricted progenitors still present some degree of plasticity[16] (**Figure 5B**).

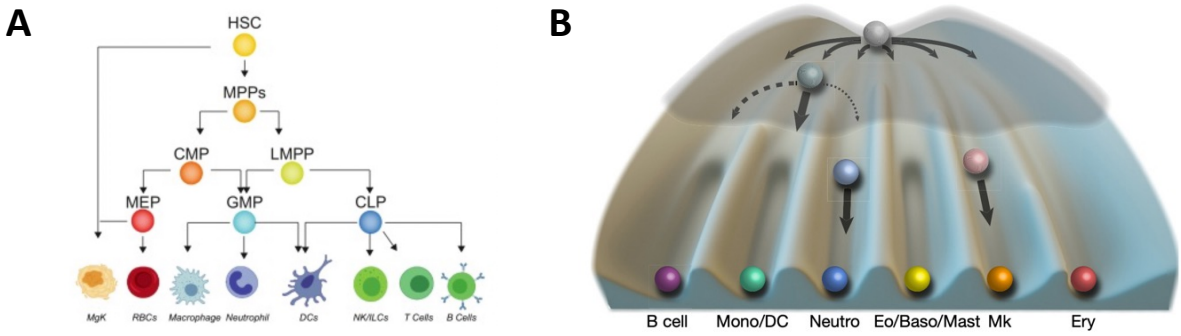


Figure 5. Schematic representation of human hematopoiesis.

(**A**) Classical representation of human hematopoiesis as a differentiation-tree model. HSC=Hematopoietic Stem Cell, MPP= Multipotent Progenitor, CMP= Common Myeloid Progenitor, LMPP= Lympho-myeloid Primed Progenitor, MEP= Megakaryocyte-erythroid Progenitor, GMP= Granulocyte-monocyte Progenitor, CLP= Common Lymphoid Progenitor. (*Adapted from Rodrigues et al. 2021*[17]). (**B**) Current representation of human hematopoiesis. HSC lineage commitment represented in a gradual manner by a continuous Waddington's landscape. Early differences in gene expression are weak and still allow the cell to pass from one state to another, while commitment is established when gene expression across lineages is too different to be overcome (*Adapted from Velten et al. 2017*[16]).

2. Erythropoiesis

Every second, the human body produces 2 million RBCs[18]. This production is ensured by a process known as erythropoiesis, i.e. the mechanism by which HSCs engage and differentiate into thousands of mature RBCs[19,20] (**Figure 5, 6**). During development, the location of erythropoiesis changes: it starts in the yolk sac in the embryo, then continues in the liver and spleen in the fetus, and after birth it occurs exclusively within the erythroblastic islands in the bone marrow. These specialized niches are composed by erythroblasts surrounding a central macrophage. The macrophage supports differentiation by interacting with erythroid precursors and by phagocyting expelled nuclei[21,22]. Erythropoiesis initiates with HSC engagement to the erythroid lineage by differentiating into more committed erythroid progenitors. The bipotential megakaryocytic-erythroid progenitor (MEP) will differentiate into burst-forming unit-erythroid (BFU-E)[18] (**Figure 6**). BFU-E are the first progenitors committed exclusively to the erythroid lineage, which in turn differentiate into Colony-Forming Unit-Erythroid (CFU-E)[18] (**Figure 6**). BFU-E and CFU-E when cultured in a semi-solid medium form large multisubunit colonies (called burst) or smaller colonies, respectively.

The second phase consists in the maturation of nucleated erythroid precursors characterized by an increasingly acidophilic cytoplasm (due to a decrease in the RNA content paralleled by an increased Hb concentration), chromatin condensation, and a decrease in cell size. During this phase, proerythroblasts and then basophilic, polychromatic, and orthochromatic erythroblasts are sequentially generated (**Figure 6**). Finally, this phase ends with the expulsion of the nucleus at the reticulocyte stage. The cells then acquire a biconcave shape by remodeling the membrane and the newly formed RBCs enter into the circulation. RBCs stay in the circulation for about 120 days before being eliminated by the spleen.

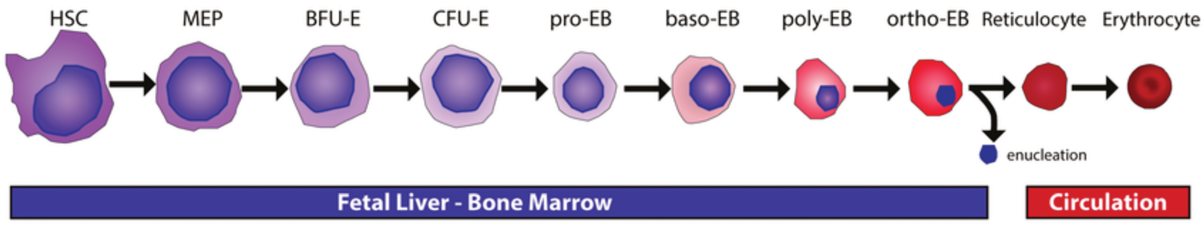


Figure 6. Overview of the erythropoiesis process.

HSCs engage to the erythroid lineage and differentiate into more committed progenitors. Then, erythroblasts terminally differentiate into enucleated reticulocytes. Reticulocytes mature into erythrocytes, which are then released to the blood stream. EB= Erythroblast. (*Adapted from Cantù et al., 2014*[23].)

C. Sickle Cell Disease

1. Epidemiology

Every year, more than 300,000 children are born with SCD worldwide[24]. Most cases are found in the sub-Saharan Africa, the Middle East, and India[25]. This distribution is partially explained by the high level of protection against malaria conferred by the sickle cell trait (i.e., carrying one mutant allele without clinically having SCD)[26]. On the other hand, migratory movements explain the current prevalence of this disease in America and the Mediterranean basin[25]. In France, SCD affects between 20,000 and 30,000 individuals[27].

2. Pathophysiology

SCD is an autosomal recessive genetic disorder. It is due to a single point mutation in *HBB*, which causes the replacement of a glutamic acid by a valine in the 6th position of the β-globin chain. This single amino acid change increases the hydrophobicity of the resulting sickle Hb (HbS) and favors the binding of 2 β^S chains from 2 adjacent HbS molecules in deoxygenated conditions (**Figure 7A**). This binding occurs thanks to the interaction of the valine in position 6 (V6) with a hydrophobic pocket formed by 3 amino acids (L88, T87, F85) from an adjacent β-globin (**Figure 7A**). Additionally, the interaction between two amino acids from α- and β-globins (H20 and E22, respectively) stabilizes lateral contacts between the Hb (**Figure 7A**). Thus, HbS

polymerization consists in the formation of long Hb fibers and cause RBCs to acquire a sickle shape in deoxygenated conditions (**Figure 7A, B**).

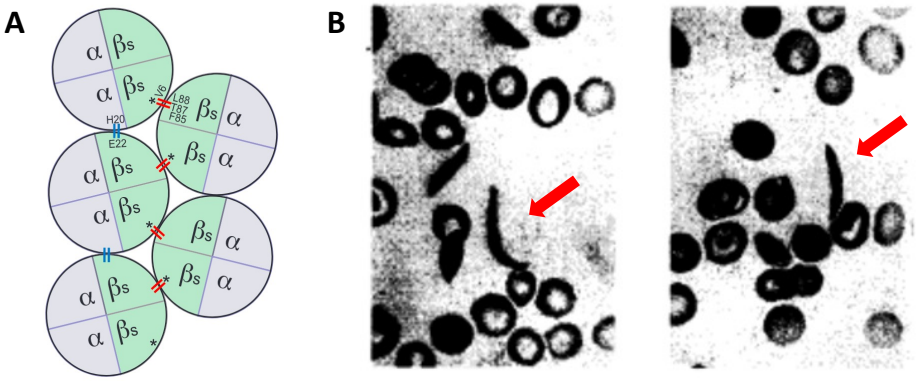


Figure 7. The polymerization of HbS under deoxygenated conditions is responsible for the deformation of SCD RBCs.

(**A**) Schematic representation of HbS polymerization and fiber formation. (*Adapted from Cavazzana et al., 2017*[2]) (**B**) Sickle shape RBCs observed in a SCD patient's blood smear. *(Adapted from Herrick et al., 2001*[28])

The autosomal recessive inheritance of SCD implies that SCD patients are homozygous for the mutated S allele, whereas heterozygous individuals are healthy carriers. However, compound heterozygous individuals with an SCD allele in combination with certain structural variants of HbA such as HbC, HbO-Arab, and HbD-Punjab or in combination with mutations on the other allele reducing the production of wild-type β-globin also present a major sickle cell syndrome[29].

The clinical manifestations are mainly due to two pathophysiological phenomena: (i) vaso-occlusive events and (ii) hemolytic anemia[29].

Vaso-occlusions occur in the microcirculation thanks to the interaction between deformed RBCs and the vascular walls (**Figure 8**). These events are promoted by hypoxia that triggers HbS polymerization and inflammation, the latter increasing cell adhesion to the endothelium[29]. Moreover, ischemia-reperfusion cycles cause oxidant stress and tissue injury[29].

On the other hand, sickling/unsickling cycles weaken RBCs and lead to hemolysis (**Figure 8**). While being responsible for severe anemia, hemolysis also contributes to the development of a progressive vasculopathy (**Figure 8**). In fact, Hb released into

the circulation generates reactive oxygen species (ROS), which are potent scavengers of nitric oxide (NO)[30]. The endothelium normally produces NO to maintain basal vasodilatation and inhibit transcription of adhesion molecules such as vascular cell adhesion molecule-1 (VCAM-1)[29]. However, free Hb participate in chronic depletion of NO and thus to endothelial dysfunction and hypercoagulability (**Figure 8**).

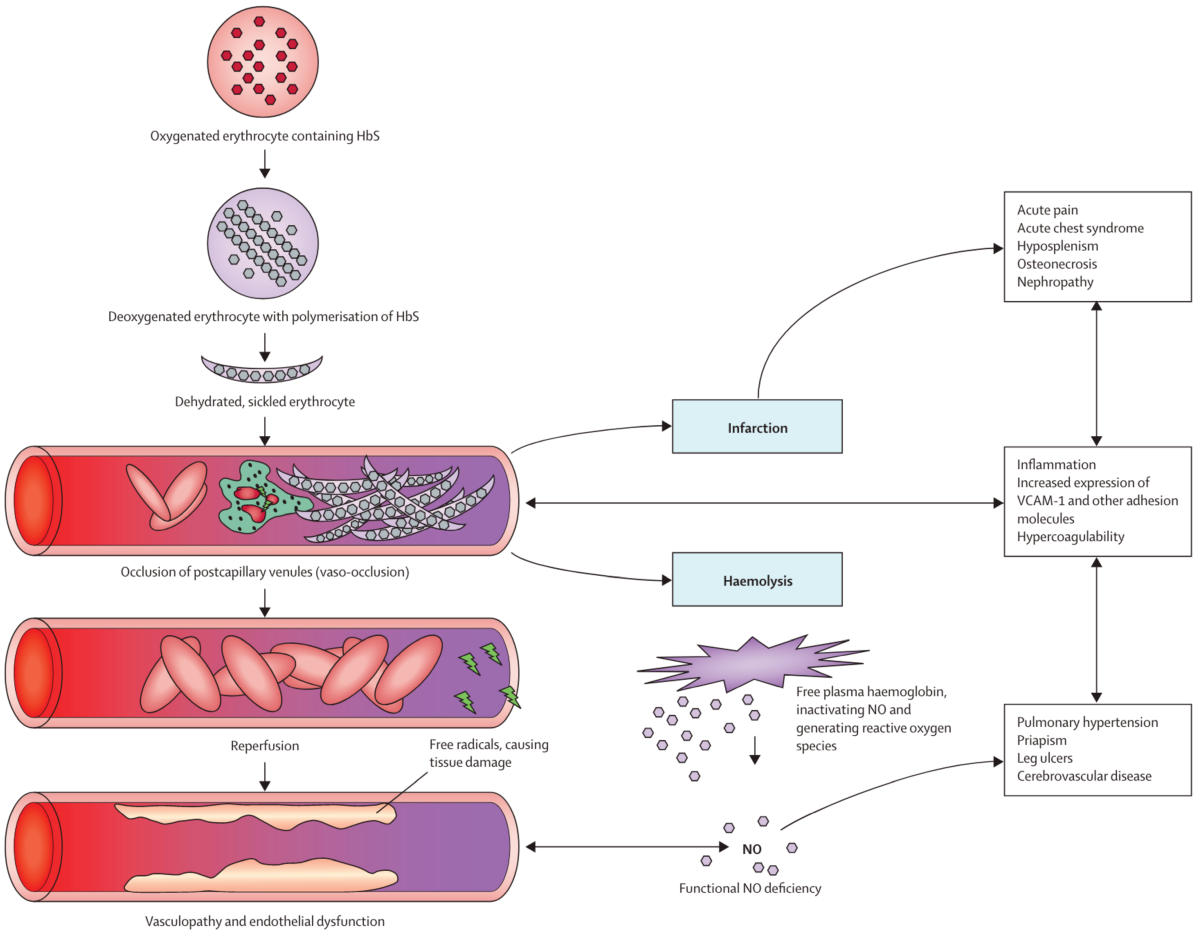


Figure 8. SCD pathophysiology.

The two key pathophysiological features of SCD are infarction and hemolysis, which in turn are responsible for vascular complications and clinical manifestations of the disease. (*Adapted from Rees et al., 2010*)[29]

3. Clinical manifestations

Sickle cell disease is characterized by acute complications such as vaso-occlusive crises and chronic hemolytic anemia (see **I.C.2 Pathophysiology**). Vaso-occlusive

crises manifest by very intense pain of sudden onset, typically affecting hands and feet in children while affecting bones and joints in adults[31]. Acute chest syndrome or stroke are serious, acute clinical manifestations. These attacks can be responsible for the necrosis of affected organs and alter their function in the long term (neurological deficits, retinopathy, cardiomyopathy, osteonecrosis, chronic renal failure, impotence, etc.). The chronic manifestations are due to the combination of the hemolytic anemia and the functional alterations of organs affected by the vaso-occlusive crises. In particular, SCD patients have an increased susceptibility to infections by encapsulated germs (e.g., septicemia, pneumonia or meningitis) due to functional asplenia caused by repeated splenic infarctions. This is one of the most important causes of illness and death in young SCD patients[29]. Moreover, vascular-endothelial dysfunction leads to chronic vascular complications such as cerebrovascular disease and pulmonary hypertension. Importantly, the clinical manifestations of SCD are extremely variable from one subject to another. As the major determinant of the disease severity is the rate and extent of HbS polymerization, the SCD phenotype can be alleviated by the coinheritance of mutations that either increase HbF levels (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**) or reduce α-globin expression that leads to reduced HbS concentration in RBCs[32].

4. Treatments

The treatment of SCD is based on 2 approaches: (i) management of acute complications, and (ii) disease-modifying treatments.

Acute complications are mainly represented by vaso-occlusive pain crises. Apart from treating the cause (rewarming, hydration, oxygen therapy, antibiotic therapy, etc.), these very painful attacks must be treated with rapid and effective analgesics. Severe pain requires hospitalization and intravenous administration of morphine derivatives. Emergency hospital treatment is essential for vaso-occlusive crises with a high risk of complications (e.g., acute chest syndrome or stroke). These complications, in addition to specific treatments, may require blood transfusion or even a transfusion exchange, i.e. partial removal of patients' RBCs in parallel with

normal RBC transfusion, to improve tissue oxygenation and reduce the concentration of circulating HbS.

Prevention of these events in the long term may require the implementation of a transfusion program depending on the frequency of complications. Transfusions need to be combined with iron chelation therapy to limit the side effects of post-transfusion iron overload. In addition, as increased erythropoiesis increases the risk of folate deficiency, daily folic acid supplementation is recommended[33]. Finally, to prevent the occurrence of serious infectious episodes, antibiotic prophylaxis is implemented during childhood, and vaccination recommendations are extended[33].

Different symptomatic drugs can help to reduce the frequency of SCD symptoms. The first drug to be approved was hydroxyurea. Although primarily known as an antineoplastic agent, this treatment is effective in reducing the acute clinical complications such as painful vaso-occlusive events and acute chest syndrome, as well as the need for blood transfusions and hospitalization[34].

How hydroxyurea improves the SCD phenotype remains uncertain but several mechanisms have been proposed. Beneficial higher HbF levels are observed under treatment and are probably due to the “stress erythropoiesis” induced by temporary hematopoiesis arrest caused by the antimetabolite action of hydroxyurea[34]. It was also suggested that it could favor ineffective erythropoiesis and positively select cells expressing high levels of HbF[35]. Moreover, cytotoxicity also reduces marrow production of white blood cells, whose high peripheral counts have been associated with increased morbidity and mortality in SCD due to their role in vaso-occlusion[36]. Finally, the hydroxyurea molecule ($CH_4N_2O_2$) contains a NO that can be released in the circulation and counteract the depletion of NO induced by hemolysis (see **I.C.2 Pathophysiology**)[37]. However, this treatment is not without potential serious adverse effects. The most common short-term adverse event is reversible myelosuppression[36]. Moreover, hydroxyurea affects DNA repair mechanisms and might be responsible for genotoxicity and consecutive malignancy. While current reports highlight the relative safety of long-term hydroxyurea usage in SCD patients, treatment initiation must be done after careful risk and benefit discussion[36].

More recently, two drugs were authorized for use in the US and European Union. The first one, crizanlizumab, is a monoclonal antibody which inhibits P-selectin, a cell adhesion molecule found on the surface of activated endothelial cells. It reduces cell adhesion to the endothelium surface and consequently limits the occurrence of vaso-occlusive crisis[38]. On the other hand, voxelotor is a Hb polymerization inhibitor, which binds reversibly to HbS and stabilizes the oxygenated Hb state. This drug was shown to significantly increase Hb levels and reduce hemolysis markers[39].

In the US, L-glutamine was recently approved by the Food and Drug Administration (FDA). This drug acts by increasing the oxidation-reduction potential of RBCs leading to a decrease in the cell adhesion to the endothelium and finally to a decrease in the occurrence of vaso-occlusive events[39].

To date, the only definitive treatment is allogeneic bone marrow transplantation from a healthy HLA-compatible donor (see **I.G.1 Gene therapy: rationale and principle**). Gene therapy approaches are also under clinical development (see **I.G Gene therapy strategies for the treatments of β-hemoglobinopathies**).

D. β-thalassemia

1. Epidemiology

Due to its protective effect against malaria[40], β-thalassemia is mainly found in populations still or previously affected by *Plasmodium*, i.e. in the sub-Saharan Africa, the Middle East, the Mediterranean basin, and the southeast Asia. More recent migratory movements explain the presence of β-thalassemia in America and Europe[41]. Overall, each year, approximately 60,000 children are born with symptomatic forms of β-thalassemia[42]. In France, β-thalassemia affects more than 700 patients[43].

2. Pathophysiology

β-thalassemia is an autosomal recessive disorder. It can be caused by more than 350 mutations responsible for decreased or absent production of β-globin; these are

referred to as β^+ or β^0 mutations, respectively[41]. Most of the mutations consist of substitutions of one nucleotide, or small insertions/deletions. These mutations result in an imbalance in the production of β- and α-chains. As a consequence, two major factors contribute to the pathophysiology of β-thalassemia: (i) the lack of β-chains and (ii) the excess of α-chains. First, the lack of β-chains leads to low levels of HbA. Thus, resulting hypochromic and microcytic RBCs are responsible for tissue hypoxia poorly compensated by marrow expansion and extramedullary hematopoiesis[41] (**Figure 9**). On the other hand, the excess of unstable free α-chains generates cytotoxic ROS and α-precipitates[18,44]. The co-inheritance of extra *HBA* genes can increase disease severity, thus highlighting the importance of the excess of α-globin chains in the disease pathophysiology[45]. The excess of free α-globin is normally eliminated by the ubiquitin-proteasome system and autophagy[46]. However, the capacity of these mechanisms is exceeded in β-thalassemia, resulting in α-globin chain accumulation that impairs the maturation and viability of erythroid precursors thus resulting in ineffective erythropoiesis and hemolysis of circulating RBCs[44] (**Figure 9**). Moreover, α-globin in excess binds to the cytoplasmic side of the RBC membrane, leading to membrane alterations[47]. Thus, the altered RBC membrane together with hemolysis and the consequent platelet activation leads to a hypercoagulable state[48] (**Figure 9**). Finally, ineffective erythropoiesis and hypoxia are responsible for increased iron intestinal absorption. As the human body cannot excrete iron in excess, it generates primary iron overload[48] (**Figure 9**).

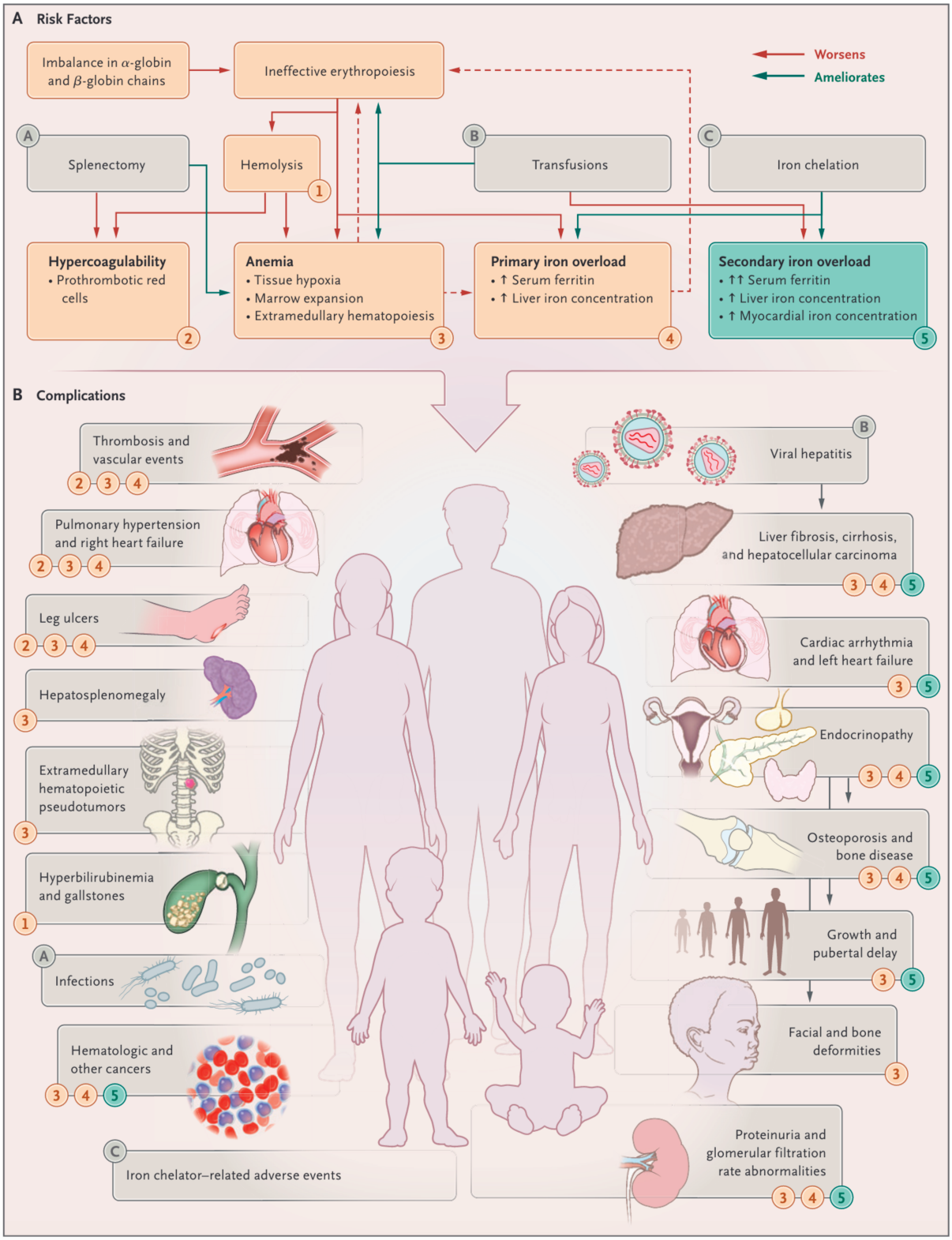


Figure 9. β-thalassemia pathophysiology and clinical manifestations.

β-thalassemia symptoms associated with the main pathophysiological characteristics or treatment-related complications. Circled numbers and letters link clinical manifestations with associated risk factors. (*Adapted from Taher et al. 2021*[41])

3. Clinical manifestations

β-thalassemic disorders are classified into 3 categories:
(i) β-thalassemia major typically includes subjects presenting a β^0/β^0 genotype or a genotype of equivalent severity involving severe β^+ mutations, with dominant forms being rare[49]. Patients present a profound microcytic hemolytic anemia and transfusion dependence.
(ii) β-thalassemia intermedia is a very heterogeneous group in terms of clinical manifestation and genotype. This group is typically represented by β^+/β^+ or β^0/β^+ subjects. Patients present clinically mild anemia and may require occasional to intermittent transfusions.
(iii) β-thalassemia minor, or so called β-thalassemia trait, includes heterozygous carriers of a β-thalassemia mutation. The subjects are clinically asymptomatic but have hematological features[50].

Phenotype heterogeneity reflects the molecular and allelic heterogeneity of this disorder, which leads to different levels of α- and β-chain production imbalance. In addition, the phenotype may be influenced by the presence of other mutations affecting the expression of α-chains (reducing the pool of α-chains and their precipitation) or HbF levels (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**).

In the case of β-thalassemia major, symptoms appear early during the first months of life whereas β-thalassemia intermedia can be manifested later during life. Clinical manifestations are a consequence of the main pathophysiological characteristics. Marrow expansion and extramedullary erythropoiesis cause hepatosplenomegaly as well as facial and bone deformities but can also lead to extramedullary hematopoietic pseudo-tumors (**Figure 9**). In addition, the excessive destruction of RBCs further contributes to splenomegaly, and bilirubin released from hemolytic RBCs can cause jaundice and gallstones (**Figure 9**). Primary iron overload is responsible for cardiac arrhythmia, liver disease, endocrinopathy, and osteoporosis (**Figure 9**). Anemia by itself leads to severe fatigue and growth retardation, but secondary tissue hypoxia also favors cutaneous complications such as leg ulcers (**Figure 9**). Finally, the

hypercoagulability state favors thrombotic events, pulmonary hypertension, and right heart failure (**Figure 9**).

4. Treatments

In major forms, regular RBC transfusions aim to correct anemia and improve erythropoiesis. An iron-chelating treatment must be associated in order to avoid the cardiac, hepatic, and endocrine consequences of iron overload due to long-term transfusions[51].

In intermediate forms, the treatment aims at counteracting the present signs of anemia or dyserythropoiesis. Splenectomy is used to limit peripheral destruction of RBCs but, in turn, increases the risk of infections and the risk of thromboembolism.

Luspatercept, an activin receptor ligand trap, acts as an erythroid maturation agent. In fact, by competing with extracellular domains of TGF-β-like molecule receptors, it inhibits downstream signaling pathways involved in erythroid maturation inhibition[52]. Thus, Luspatercept, by restoring normal erythropoietic differentiation, improves anemia and reduces transfusion requirements. It was approved by the FDA and the EMA in 2019 and 2022, respectively[52].

To date, the only curative treatment is allogeneic bone marrow transplantation from a healthy HLA-compatible donor (see **I.G.1 Gene therapy: rationale and principle**). Moreover, Zynteglo® received approval in Europe and in the US as the first gene therapy product indicated for transfusion-dependent β-thalassemia. This product leads to β-globin production restoration by integrating copies of the *HBB* gene through lentiviral transduction in the genome of patient HSCs (see **I.G.2 Lentivirus-based strategies**).

E. CRISPR/Cas9

Genome editing consists in the replacement, the deletion or the insertion of a DNA sequence in the genome of a living organism. The efficiency of genome editing was raised by inducing double-strand breaks (DSBs) at the target locus within the

genome. Engineered nucleases such as zinc-finger nucleases (ZFN) and transcription-activator-like effector nucleases (TALEN) were among the first tools used to induce DSBs at a specific locus within the genome[53]. However, the requirement for substantial protein engineering and limited efficiencies restricted their applications. The simplicity of the Clustered Regularly Interspaced Short Palindromic Repeats and the CRISPR-associated protein 9 (CRISPR/Cas9) design together with its high efficiency has significantly facilitated the wide use of genome editing.

1. CRISPR/Cas9 in prokaryotic organisms

CRISPR/Cas is an adaptive defense system used by prokaryotes to cleave non-self nucleic acid sequences[54]. The CRISPR/Cas system consists of the CRISPR locus and the Cas genes. Different types of CRISPR systems and Cas proteins have been described according to their genetic structure. However type II CRISPR system and its Cas9 effector protein are the mostly used for gene editing[55,56].

Upon infection, the infected prokaryotic cell equipped with the CRISPR/Cas system integrates into its genome small DNA fragments from the invader's genome. These sequences are inserted at the CRISPR locus where they are interspersed between short repeats (**Figure 10**: *Acquisition*). This locus is then transcribed into a long pre-CRISPR RNA (pre-crRNA) that is cleaved into small crRNAs (**Figure 10**: *crRNA biogenesis*). Each crRNA consists of a repeat sequence and a sequence complementary to the known foreign DNA, called a spacer. The repeat sequence allows the recruitment by sequence complementarity of a trans-activator RNA (tracrRNA), which in turn, thanks to its complex secondary structure, allows the recruitment of the Cas9 protein and stabilizes its interaction with the crRNA-tracrRNA complex[57] (**Figure 10**). The Cas9 protein is an endonuclease whose activity is guided by the crRNA spacer sequence and induces a DSB at the recognized complementary sequence, the protospacer. This system allows the destruction of foreign DNA in case of reinfection by a known pathogen (**Figure 10**: *Interference*).

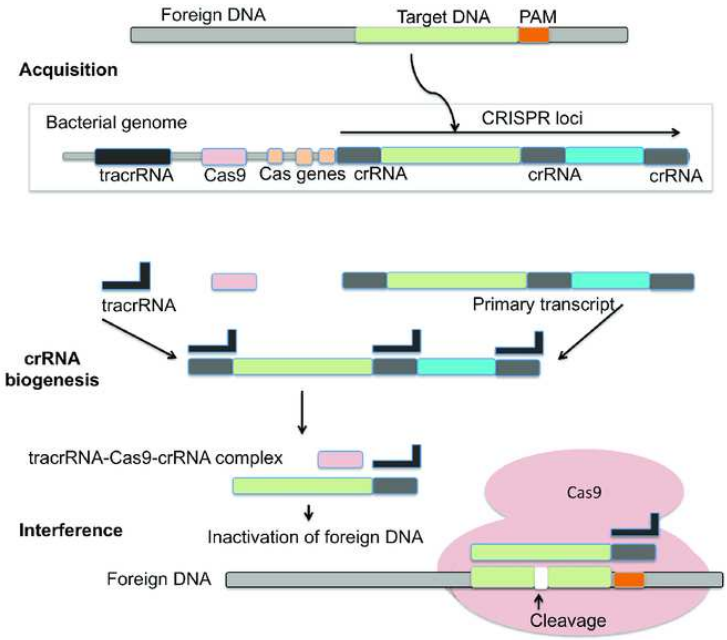


Figure 10. Schematic representation of the CRISPR/Cas9 system mechanism in prokaryotes.

Upon infection, short fragments of DNA from the invader are copied into the CRISPR locus. Transcribed crRNAs from sequences encoded at this locus guide the Cas9 endonuclease that induces DSBs in the invader DNA upon reinfection. (*Adapted from Arora et al., 2017*[58])

In order for this system to work properly, it is important that Cas9 distinguishes between protospacer sequences in heterologous DNA and those that have been integrated into the genome, to specifically cut the invading DNA. Therefore, Cas9 first recognizes a specific protospacer adjacent motif (PAM) present right after the protospacer sequence only in the foreign DNA (**Figure 10**, **Figure 11**). For instance, *Sp*Cas9, the most frequently used Cas9 protein from *Streptococcus pyogenes*, recognizes the -NGG PAM. If this sequence is present, then the upstream DNA is unwound looking for complementarity with the crRNA spacer sequence (**Figure 11**). If complementarity is found, Cas9 generates a DSB three nucleotides upstream of the PAM (**Figure 11**).

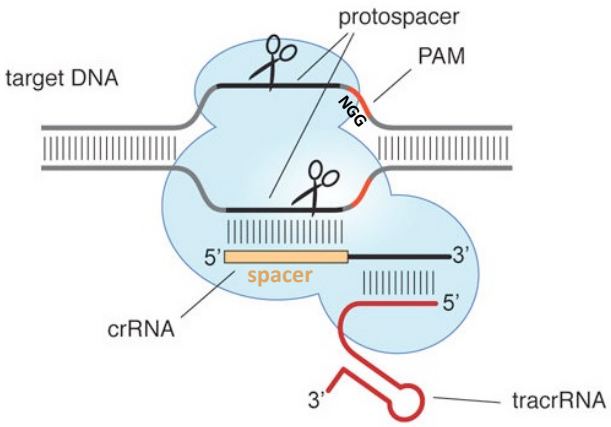


Figure 11. Recognition of the protospacer sequence by the Cas9 protein guided by the crRNA-tracrRNA duplex.

The complex is formed by a structure of 2 non-coding RNAs: the crRNA carrying the spacer sequence complementary to the target and the tracrRNA whose secondary structure stabilizes the interaction with Cas9. The dual specificity of the cleavage site is ensured by the recognition of the Cas9-specific PAM and the complementarity of the crRNA spacer sequence. (*Adapted from Jinek et al., 2012*[54].)

2. CRISPR/Cas9 as a genome editing tool

Jinek et al. described in 2012 a genome editing tool derived from the CRISPR/Cas9 system[54]. Based on the understanding of the structure formed by the tracrRNA and the crRNA, they proposed the use of a single transcript combining the spacer sequence of the crRNA with the secondary structure formed by the tracrRNA. This unique transcript, called single guide RNA (gRNA) simplifies the use of this system for genome engineering (**Figure 12**). Indeed, the only requirement for Cas9-mediated DSB at a specific locus is the adaptation of the gRNA 20 nucleotides spacer sequence in order to be complementary with the protospacer of the targeted sequence (**Figure 12**).

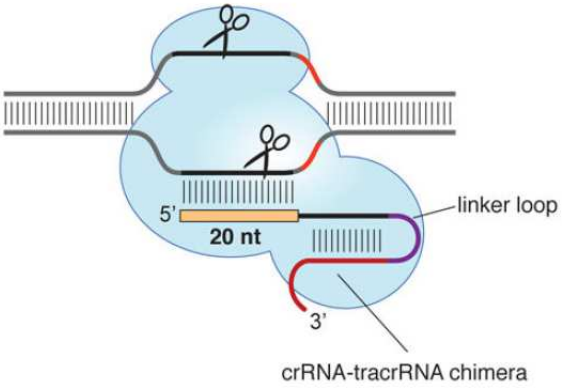


Figure 12. Recognition of the protospacer sequence by the Cas9 protein guided by a chimeric RNA.

The gRNA contains a spacer sequence complementary to the target at the 5' end, followed by the secondary structure characteristic of the crRNA-tracrRNA complex at the 3' end. It allows the Cas9 protein to be guided to the targeted cleavage site. (*From Jinek et al., 2012*[54])

Once Cas9 and gRNA have been delivered into the cell, the DSB generated by Cas9 can be repaired by different mechanisms depending on the cell type and the editing conditions. Thus, the type of modification induced at the cleavage site depends on the DSB repair mechanism used by the cell (**Figure 13**). DSB repair by non-homologous end joining (NHEJ) results in small insertions and deletions (InDel) around the target site, and can be exploited to inactivate genes or regulatory sequences (**Figure 13**). On the other hand, if two homologous sequences are present on both sides of the cut site, then microhomology-mediated end joining (MMEJ) can take place and leads to specific deletions (**Figure 13**). Finally, addition of an exogenous DNA template homologous to the target sequence triggers homology-directed repair (HDR) and thus makes it possible to precisely replace one sequence with another one. Although HDR seems to be an interesting method to correct the mutations responsible for β-hemoglobinopathies, its efficiency is limited in quiescent cells such as HSCs[59].

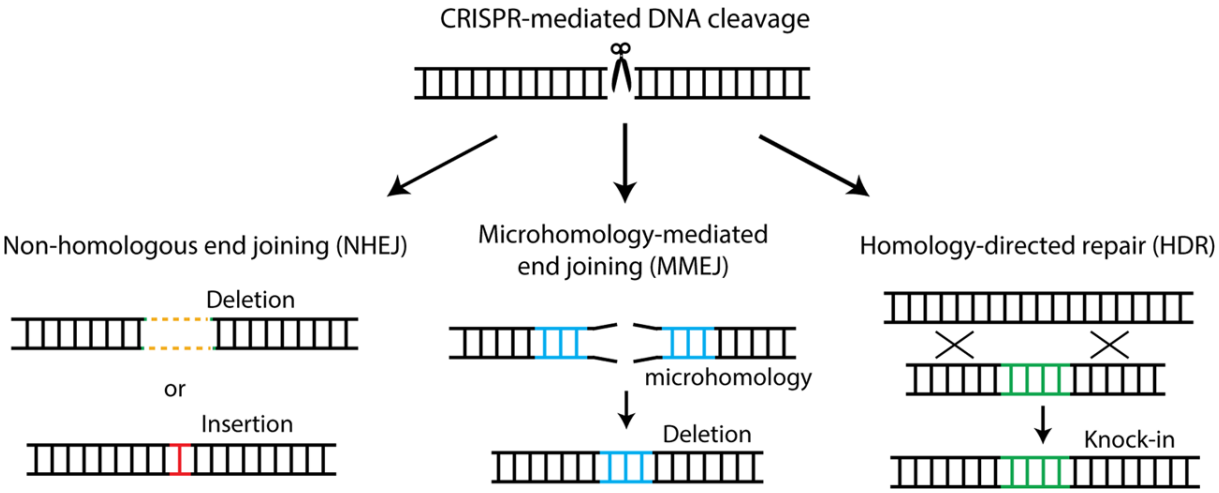


Figure 13. Sequence modifications induced by the different repair mechanisms of CRISPR/Cas9-induced DSB.

Depending on the repair mechanism involved, a CRISPR/Cas9-induced DSB can result in insertions, deletions, or allow the replacement of the targeted sequence with a homologous sequence that includes the desired modification. (*From Jang et al., 2020*[60].)

3. Safety of CRISPR/Cas9 usage

Several concerns are raised by the CRISPR/Cas9 usage. In particular, its activity can lead to editing byproducts at on- and off-target sites, but also to a DNA Damage Response (DDR) in targeted cells.

a) Editing byproducts at the on-target site

The on-target genotoxic risk is raised by unwanted genomics modifications occurring at the targeted site. These modifications vary from a few base pair unwanted edits to large chromosomal rearrangements. The most frequent type of these undesired events consists of the involvement of an unwanted DNA repair mechanism. In particular, genome editing strategies aiming to correct a gene sequence through HDR suffer from important editing rates resulting from NHEJ. Thus, instead of being corrected, the targeted gene can be inactivated. On the other hand, it was demonstrated that CRISPR/Cas9 can also induce large deletions, chromosomal loss, or chromothripsis (extensive chromosome rearrangements restricted to one or a few chromosomes known to be involved in cancer)[61–64].

b) Editing byproducts at the off-target sites

Off-target activity results from aspecific binding and DSB generation in non-targeted genomic loci. These events are mainly due to sequence similarities between the targeted locus and the off-target one. In fact, gRNA hybridization can tolerate a few mismatches, especially in the 5' part of the gRNA[65]. Off-target DSBs raise safety concerns as not only they might be deleterious by themselves if they occur in a sensitive site (e.g. a gene or a regulatory sequence) but they can also trigger complex chromosomal rearrangements, such as translocations between the on-target and the off-target sites[61,66].

c) DNA Damage Response

CRISPR/Cas9 treatment triggers the activation of DDR, a pathway that aims to maintain genome stability in response to DNA damage[67]. In particular, DSBs can activate p53-dependant cell cycle arrest. Activation of this pathway affects HSC functionality and leads to a reduced repopulating potential[68,69]. Thus, edited cell populations might be counter selected due to DDR activation. However, under these circumstances, edited cells presenting a deficient p53 pathway might present a selective advantage. Given the tumor suppressor function of p53, this growth bias is problematic as it could potentially support tumorigenesis within the edited cell population[68,70,71].

F. Base editing

1. Development of base editors

Base editing allows the insertion of point mutations at a specific locus without the need for a DSB and the requirement of an exogenous DNA template, therefore it represents a safe and less complex strategy to correct gene mutations. Base editors were developed by combining a catalytically inactivated Cas9 nuclease with a deaminase (**Figure 14**). Inactivation of *Sp*Cas9 by insertion of two mutations (D10A and H840A) generates the dead Cas9 (dCas9)[72]. While preserving its ability to recognize and bind to a specific sequence in the genome, the dCas9 is devoid of nuclease activity.

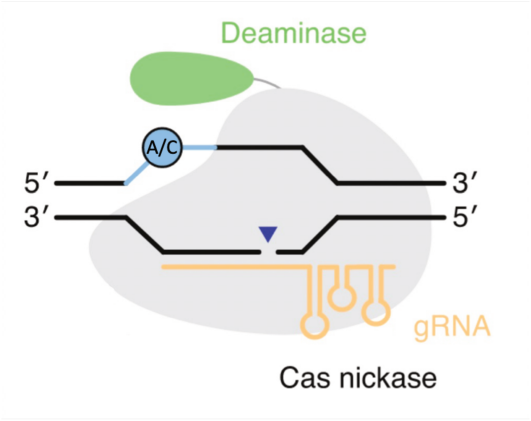


Figure 14. Adenine and cytosine base editors.

The latest generation of base editors consists of two subunits: a Cas9 nickase (in gray) linked to a deaminase (in green). The Cas9, guided by the gRNA (orange), allows the formation of a R-loop at the target sequence and thus exposes the non-complementary strand to deaminase activity. (*From Anzalone et al. 2020*[73])

Depending on the type of deaminase, two types of base editors can be distinguished (**Figure 14, 15**). Adenine base editors (ABEs), contain an enzyme that replaces an adenine (A) with a guanine (G) via intermediate conversion to an inosine (I)[74] (**Figure 15A**). Cytosine base editors (CBEs) contain a cytosine deaminase that allows the replacement of a cytosine (C) with a thymine (T) via intermediate conversion to an uracil (U)[72] (**Figure 15B**).

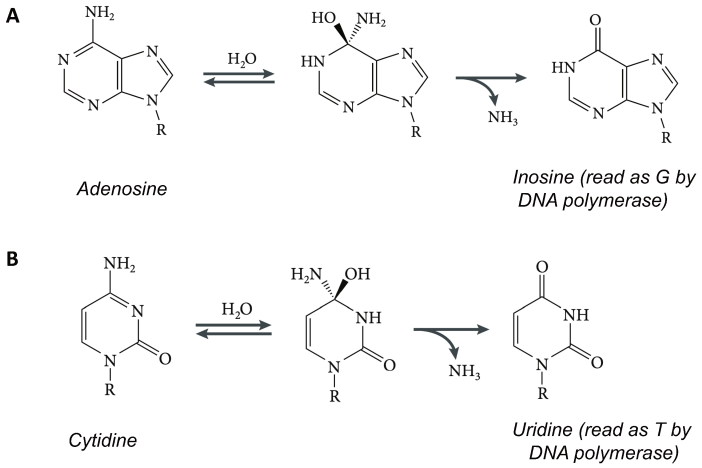


Figure 15. Chemical reactions catalyzed by base editors.

(**A**) Adenosine deamination generates inosine, which is recognized as a guanosine by the DNA polymerase. (**B**) Cytidine deamination generates uridine, which is recognized as a thymidine by the DNA polymerase. (*Adapted from Rees et al., 2018*[75])

The deaminase mainly used in CBEs is the rat deaminase rAPOBEC1, which was selected for its high cytosine deaminase activity[72]. Regarding ABEs, the absence of DNA adenine deaminase prompted Gaudelli et al. to engineer the TadA *E.coli* tRNA adenine deaminase to target and modify DNA (hereinafter referred to as TadA*). As TadA works naturally as a dimer, better efficiencies were obtained when fusing the Cas9 with the heterodimer TadA*-TadA instead of fusing it only with TadA*[74].

The pairing of the gRNA to the targeted strand results in the formation of a DNA:RNA hybrid complex and exposes the non-complementary strand in a single-stranded form (**Figure 14, 16**). This structure, called the R-loop, allows the deaminase to target the non-complementary strand and deaminate accessible bases (**Figure 16**: *steps 2 to 3*). Editing of the non-complementary strand induces a mismatch. This situation can be resolved by the DNA mismatch repair (MMR) mechanism either by restoring the modified base or by replacing the base on the non-deaminated strand with a compatible one[76]. In this case, U having the same pairing properties as T, it will be paired with an A, whereas I will be paired with a C (**Figure 15**, **16**: *steps 3 to 4*). Finally, the U and I intermediates are replaced by DNA repair or during replication with a T or a G respectively[72,74]. However, as the presence of U and I intermediates in DNA results from a premutagenic event, these bases might trigger DNA repair pathways before being replaced by T and G, respectively. Thus, U can be excised by uracil nuclear DNA N-glycosylase (UNG2) and then be reverted to the original sequence by the base excision repair (BER) pathway[76]. On the other hand, the I intermediate can be targeted by the alkyl adenine DNA glycosylase (AAG) and replaced by the BER mechanism[76]. Whereas CBEs editing outcome is highly improved in terms of efficiency and purity by uracil glycosylase inhibitor (UGI) fusion to CBEs, inhibition of the equivalent enzyme involved in I excision did not improve ABEs editing outcomes, suggesting that I excision is less efficient and does not limit the efficacy of ABEs[73,74,76].

The initially-developed system was still limited by the fact that the DNA repair machinery could solve the mismatch by restoring the modified base, or in other words by using the targeted non-deaminated strand as a template for resynthesizing the opposite strand. Thus, Komor et al. devised a way to promote base replacement on the opposite strand rather than restoration of the edited base[72]. Reversion of the

H840A mutation from the dCas9 generates the Cas9 nickase (nCas9), which retains its nuclease activity only on the gRNA-paired strand. Thus, nicking the unedited strand triggers DNA repair machinery on this side, and increases the efficiency of base editing (**Figure 14, 16:** *steps 2 to 4*).

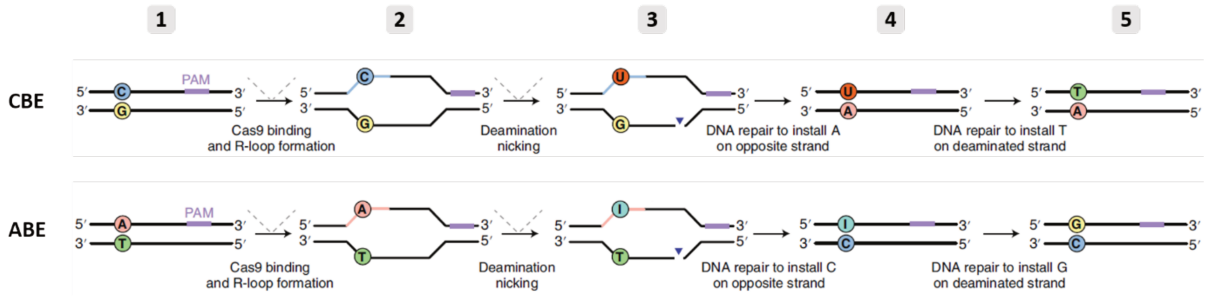


Figure 16. Sequence of events leading to the insertion of a mutation by base editing.

The non-targeted strand (containing the PAM) is exposed to deaminase activity (steps 1 to 2). Depending on the type of base editor, the enzyme allows the deamination of C to U or A to I (steps 2 to 3). The opposite unmodified strand is cleaved by nCas9, thus favoring its reparation to solve the mismatch (steps 3 to 4). U are replaced by T and I by G by the intervention of DNA repair enzymes and upon DNA replication. (*From Anzalone et al., 2020*[73])

2. Genome editing with base editors: requirements and optimizations

Efficient base editing relies on the combination of the gRNA and the base editor used. Design of the gRNA needs to meet several requirements. First, its target sequence must be directly followed by an appropriate PAM sequence (**Figure 11**). Second, the targeted base should be placed within the “editing window”, which corresponds to the positions of the protospacer that will be accessible to the deaminase (**Figure 17**). Typically, for canonical ABEs, the editing window corresponds to positions 4-7 of the protospacer, whereas it corresponds to positions 4-8 for canonical CBEs[75] (**Figure 17**). Efficient base editing requires also the choice of the appropriate base editor.

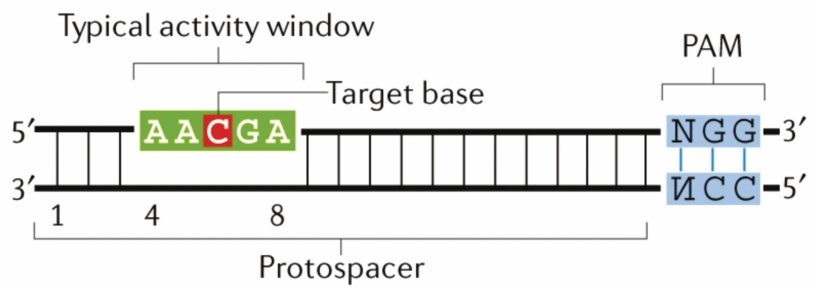


Figure 17. Schematic representation of the protospacer and the base editing activity window.

Typical base editing window corresponds to the protospacer positions 4-7 for canonical ABEs, or positions 4-8 for canonical CBEs. (*Adapted from Rees et al., 2018*[75])

As usage of the first base editors was restricted by canonical PAM availability near the targeted base, several Cas9 variants were developed in order to target non -NGG PAMs[77–80]. However, the targeting scope of engineered Cas9 was still largely restricted to PAM sequences containing Gs. Therefore, Miller et al. developed three Cas9 variants recognizing the non-NGG PAMs -NRCH, -NRTH, and -NRRH (R=A or G, H=C,T or A) thus increasing substantially the fraction of accessible genomic sites[80]. Moreover, Walton et al. engineered *Sp*Cas9 to relax its PAM requirement and efficiently target NRN and less efficiently NYN PAMs (R=A,G and Y=C,T) thus creating a virtually PAM-less Cas9 called SpRY[81].

While CBEs are broadly compatible with alternative Cas9 orthologs, ABEs suffer from substantially reduced efficiencies, limiting their broad usage[82]. In order to overcome this limitation, Richter et al. developed a highly processive adenosine deaminase, by selecting 8 mutations in the TadA domain[82]. The ABE8e enzyme containing the evolved TadA-8e domain showed substantially increased base editing efficiencies but also broadened the editing window, while maintaining high rates of editing with alternative Cas9 variants and orthologs[82].

Moreover, Zhang et al. obtained higher editing rates and expanded editing windows of a CBE variant by fusing the single-strand DNA-binding domain (ssDBD) Rad51 in between the Cas9 and the deaminase[83] (**Figure 18**). Likewise, Zafra et al. obtained similar results by fusing two extra nuclear localization signals (NLS) within the linker sequence of a CBE variant[84]. Mechanisms underlying this higher editing efficiency are not elucidated yet and might result depending on the construct either from the

higher binding of the base editor to the ssDNA, higher nuclear localization or to the fact that a longer linker between Cas9n and the deaminase (because of the presence of the ssDBD) sets free the deaminase and facilitates its interaction with the substrate[83] (**Figure 18**).

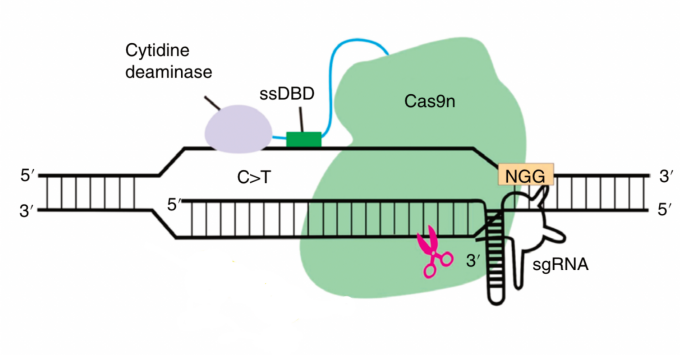


Figure 18. Single-strand DNA-binding domain addition increases CBE activity.

Addition of a single-strand DNA-binding domain within the linker fusing the Cas9 to the deaminase was described as a strategy to increase base editing efficiencies and modulate the base editing window of a CBE. (*Adapted from Zhang et al., 2020*[83])

3. Safety of the base editing procedure

a) Editing byproducts at the on-target site

Three types of undesired events can happen at the on-target site: transversion mutations at the target nucleotide, bystander edits, and InDels.

Transversion mutations at the target nucleotide consist of A>non-G or C>non-T edits. While these types of edits were largely described with CBEs, ABEs demonstrated high product purity[74]. It can be explained by the fact that transversion edits probably arise from error-prone repair of abasic sites after DNA glycosylase base excision[73] (see **I.F.1 Development of base editors**). However, the lower activity of this mechanism observed in the frame of ABE editing (see **I.F.1 Development of base editors**), might explain the higher purity of ABE outcomes.

Bystander edits consist in the deamination of non-targeted A or C found in the editing window (**Figure 19**). In this situation, the effect of bystander edits must be evaluated, and the potential biological effects should be considered before clinical application. In

particular, if the target sequence is a protein-coding region, this edit might lead to a non-silent mutation and impact the protein structure and function. Moreover, it was also demonstrated that "silent" mutations could disturb mRNA stability and translation[85,86]. Of note, even though ABEs were described to result in very high product purity, some papers described low C>T and C>G editing as bystander edits in human cells[87,88]. In general, bystander edits can be limited by base editors with a narrowed editing window or a context-specific activity[89–91].

Finally, base editors contain a nCas9, which simply nicks the targeted strand thus generating single strand breaks (SSB). However, possible recognition and excision of the deaminated intermediate through BER mechanism (see **I.F.1 Development of base editors**) might transiently generate a SSB on the opposite non-targeted strand. If this event happens before re-ligation of the nick, then it results in a DSB. In addition, if the nick persists until DNA replication, it will be treated as a DSB[73,92]. DSBs resulting from these events can be repaired by different pathways of the DNA repair machinery, similarly to DSBs generated by the Cas9 nuclease (see **I.E.2 CRISPR/Cas9 as a genome editing tool**). In particular, NHEJ and MMEJ repair pathways lead to the formation of InDels. Importantly, on-target InDels have been mainly described following CBEs usage and seem to happen to a much lower extent with ABEs, probably due to inefficient excision of inosines (see **I.F.1 Development of base editors**).

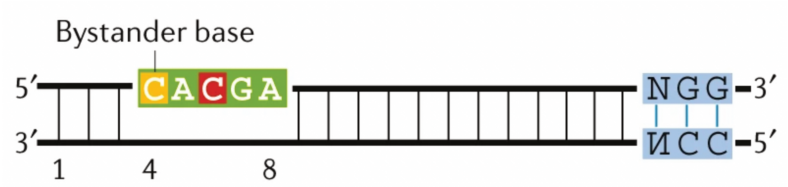


Figure 19. Schematic representation of the protospacer and bystander edits occurring within the editing window.

Bystander edits result from aspecific base editing of C (with CBE) or A (with ABE) adjacent to the target base in the editing window. Here, the target base is indicated in red, while the potential bystander edit location is indicated in yellow. Both are located within the editing window (highlighted in green). (*Adapted from Rees et al., 2018*[75])

b) Editing byproducts at the off-target sites

In addition to undesired edits generated at on-target site, deamination can also happen at off-target DNA or RNA sites. Regarding DNA off-target sites, they are either gRNA-dependent or gRNA-independent. gRNA-dependent off-targeting is due to Cas9 recognition of sequences that are similar to the on-target locus. Cas9 binding generates R-loop formation at the off-target site, thus allowing the enzyme to deaminate potentially accessible As or Cs. In order to limit these edits, high-fidelity variants can be used to increase the specificity of Cas9 binding[93,94].

gRNA-independent off-target activity is due to the non-specific interaction of the deaminase with transiently accessible DNA bases. Studies comparing the occurrence of these events with CBEs and ABEs, revealed that they happen with higher frequencies upon CBEs usage[95–98]. However, modifications in the deaminase domain or the use of alternative deaminases can diminish this type of off-target activity[95–97].

Finally, RNA off-targets can also occur in a gRNA-independent manner. Even though this effect has been described with both types of base editors, CBEs seem more prone to edit the RNA than ABEs[96,99]. RNA edits occur in both protein-coding and non-protein coding sequences[100]. Given the short half-life of RNA, these off-targets are transient but potential functional consequences must be carefully evaluated before therapeutic application of base editing approaches in humans[100]. Notably, engineering of the TadA and rAPOBEC1 domains[100–104], or even usage of alternative deaminases in CBEs[95,102,103], resulted in base editor variants with substantially reduced RNA off-target activity.

G. Gene therapy strategies for the treatments of β-hemoglobinopathies

1. Gene therapy: rationale and principle

To date, the only curative treatment for SCD and β-thalassemia is allogeneic transplantation of healthy HSCs. The best results are obtained in children receiving a graft from an HLA-matched brother or sister. For SCD treatment, the 5-year survival

rate of this treatment is around 95%[105]. However, as each sibling has only a 25% of likelihood of being HLA-identical, the application of this treatment is limited. If a matched sibling donor cannot be identified, alternative transplants can be considered from either an unrelated HLA-compatible donor or a haploidentical (HLA-semi-matched) family member. However, the use of partially match HLA transplants increases the risk of serious adverse events such as graft rejection or graft-versus-host disease[106]. To overcome these limitations, the collection of autologous HSCs and their "correction" by genetic manipulation is an interesting alternative (gene therapy; **Figure 20**).

The success of gene therapy depends on the frequency and amount of corrected HSCs injected into the patient. However, in SCD and β-thalassemia patients, HSC collection from the bone marrow does not yield a sufficient number of cells. Indeed, erythropoiesis in β-thalassemic patients is inefficient and leads to the accumulation of erythroid precursors in the bone marrow, preventing the collection of a good number of HSCs[107]. Similarly, the inflammatory bone marrow microenvironment does not allow efficient HSCs collection in SCD patients[2]. Mobilization allows HSCs to pass into the peripheral blood so that they can be collected by apheresis (**Figure 20**). Normally, HSCs are anchored in their niche by the interaction between their surface receptor CXCR4 and a chemokine constitutively produced by bone marrow stromal cells called SDF-1. HSC mobilization can be performed using drugs intervening at the level of this interaction, such as plerixafor, a CXCR4 antagonist, or G-CSF, a hematopoietic growth factor that leads to a decreased expression of SDF-1 ligand and CXCR4 receptor[108,109]. Thus, in β-thalassemia patients, it is recommended to use a combination of G-CSF and plerixafor to recover HSCs by apheresis[110]. However, in SCD patients, the use of G-CSF is contraindicated as this treatment induces severe life-threatening vaso-occlusive crises[111]. Thus, plerixafor alone has been shown to be a safe and effective means of mobilizing SCD patients HSCs, and this strategy is used in current gene therapy protocols[112,113].

After mobilization, $CD34^+$ hematopoietic stem/progenitor cells (HSPCs) are collected by apheresis (**Figure 20**). CD34 is a transmembrane glycoprotein with a role in early hematopoiesis by mediating stem cell adhesion[114]. This marker is regularly used in clinics to select and enrich HSPCs in a graft product. Then, various genetic

manipulations can be applied to the sorted cells in order to obtain a population of "corrected" HSCs ready to be re-administered to the patient.

Before receiving the graft, the patient must be treated with myeloablative chemotherapy that eliminates residual unharvested HSCs (**Figure 20**). This treatment usually consists in busulfan administration and will facilitate the engraftment of treated HSCs. After a few days of washout, the graft is injected intravenously.

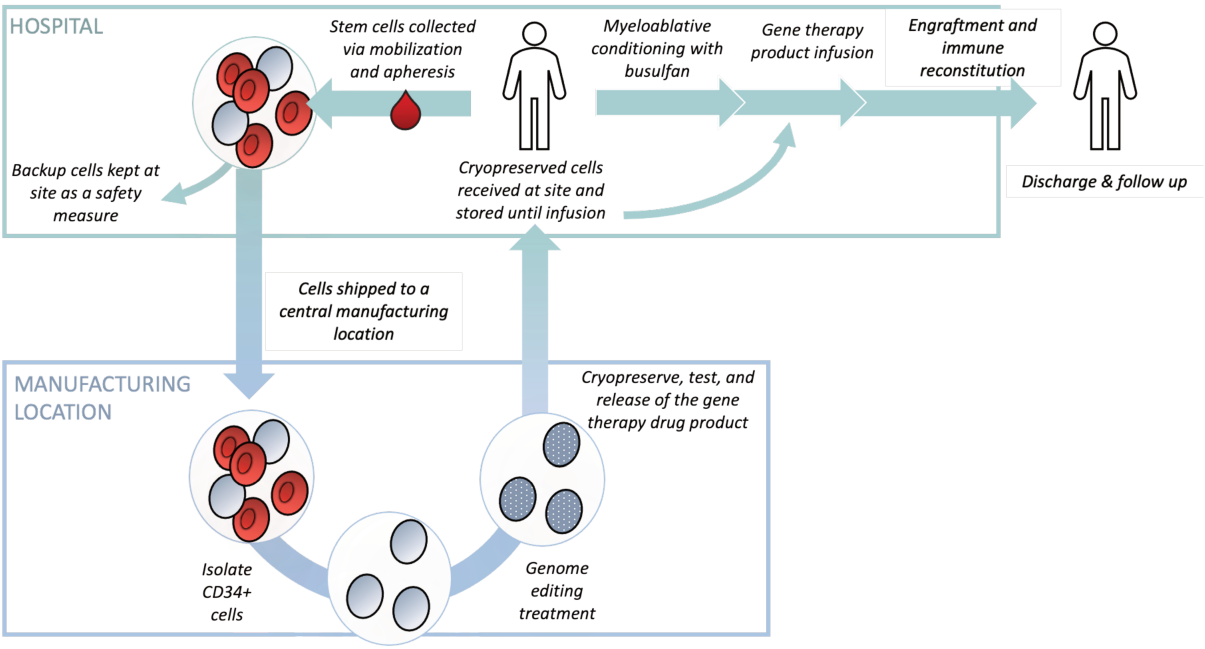


Figure 20. Gene therapy procedure by autologous transplantation of *ex vivo* genetically modified HSCs.

Patients HSCs are mobilized and collected by apheresis. After selection of the $CD34^+$ fraction enriched in progenitors and stem cells, the cells are genetically modified. The graft is cryopreserved, tested, and reinjected into the patient after myeloablative conditioning. (*Adapted from Frangoul et al. 2021*[112]).

Gene correction of the $CD34^+$ cells can rely either on gene addition (see **I.G.2 Lentivirus-based strategies**), or on genome editing (see **I.G.3 Genome editing strategies**).

2. Lentivirus-based strategies

The first gene therapy strategy used in clinics for β-hemoglobinopathies was the addition of a therapeutic gene into the HSCs genome through lentiviral vector treatment. Lentiviruses have the capacity to retro-transcribe a transgene encoded in the viral RNA and to integrate it into the DNA of transduced cells in order to achieve persistent protein expression. This strategy was first tested in 2006 in the frame of the LG001 phase 1/2 clinical trial for β-thalassemia patients. The vector used in this trial, HPV569, contained a transgene encoding $\beta^{A\text{-}T87Q}$-globin. The T87Q amino acid substitution was introduced in the transgene-derived β-globin because the Q in position 87 confers anti-polymerization properties to the γ-globin gene. Thus, this vector can be also used for SCD treatment. In the case of β-thalassemia, the therapeutic effect of this vector is obtained by restoring the α/non-α ratio and the inserted mutation makes it possible to distinguish the Hb resulting from transgene expression from endogenous Hb for patient monitoring[115]. Two severe β-thalassemia patients have been treated with this first-generation vector, one of whom became transfusion-independent[115]. This vector was subsequently optimized to increase transduction efficiency[116]. The first gene therapy successes using the optimized vector, BB305, were reported in 2017 and 2018 for SCD and β-thalassemia patients, respectively[117,118]. It is now being tested in phase 3 trials for SCD (HGB-210, NCT04293185) and β-thalassemia (HBG-207, NCT02906202 and HGB-212, NCT03207009)[119]. Similarly, a mutation was inserted in the *HBG* transgene to confer to the transgene a competitive advantage over the β^S in terms of the interaction with the α-globin. The lentivirus encoding this modified HbF, HbF^{G16D}, is currently being tested in phase 1/2 in SCD patients and shows an effect on the reduction of vaso-occlusive crises after treatment[120,121]. Another lentiviral vector, GLOBE, encoding the β-globin gene has shown efficacy in the phase 1/2 TIGET-BTHAL study (NCT02453477) in transfusion-dependent β-thalassemia patients[122]. The TIGET-BTHAL study is currently being followed up for long-term efficacy and safety assessment (NCT03275051).

Two other vectors have been developed, GLOBE1 and βAS3-FB, both containing a transgene encoding a modified β-globin AS3. This globin is called *AS* for anti-sickling, because it contains mutations generating three beneficial “anti-sickling”

amino-acids in the LV-derived β-globin chain (G16D, E22A, T87Q). These vectors are currently being tested for SCD treatment in phase 1/2 clinical trials (NCT03964792 and NCT02247843).

Finally, a lentiviral vector encoding a microRNA that decreases the expression of BCL11A, a master repressor of HbF (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**), has been shown to be effective in reactivating HbF expression and reducing the symptoms of SCD in 6 patients (NCT03282656)[123].

Despite very encouraging results, gene addition by viral vectors faces some limitations. Indeed, the more severe is the phenotype, the more variable is the response. For instance, it seems that Hb resulting from transgene expression is insufficient in severe β^0/β^0 β-thalassemia patients and that they would require the incorporation of a greater number of vector copies per cell to achieve the same results as those obtained in non-β^0/β^0 patients[118]. Efforts are focused on optimizing transduction efficiencies to ensure a large supply of modified cells and high levels of transgene expression to achieve more consistent results. However, increasing transgene integration events theoretically increases the risk of genotoxicity. In fact, integrating viruses, by inserting themselves into the host genome, can theoretically trigger oncogenesis by upregulating proto-oncogenes or disrupting onco-suppressor genes. This adverse event is called insertional mutagenesis.

3. Genome editing strategies

Today, new genome editing strategies offer the possibility to correct the mutations responsible for β-hemoglobinopathies and restore endogenous *HBB* expression, or to induce high levels of endogenous *HBG* expression. While lentiviruses cannot include the entire *HBB* LCR regulatory sequence because of packaging limit, genome editing-mediated re-expression of the endogenous *HBG*/*HBB* genes has the advantage of benefiting from the physiological genetic environment of the gene and thus allowing a higher level of therapeutic globin expression. Moreover, these strategies do not require the insertion of exogenous DNA, therefore avoiding the risk of insertional mutagenesis.

Genome editing in HSCs requires the delivery of the editing machinery to the target cell population. Importantly, for clinical application, the delivery method must allow a transient expression of the enzyme in order to decrease off-target DNA and RNA modifications and the potential immune response *in vivo* to the editing machinery (e.g., mRNA, or pre-complexed ribonucleoprotein (RNP))[93,101].

For genome editing, *ex vivo* approaches allow the direct administration of the enzyme/gRNA into HSCs through electroporation. This technique exerts a strong electric field onto the cell membrane and increases its permeability. It has the advantage of being very efficient in HSPCs and allowing transient delivery of large molecules. The most straightforward way to deliver the editing machinery in target cells is to transfect pre-complexed RNP. While this delivery method has shown its efficiency with the Cas9 nuclease, base editor protein production remains challenging[124]. Thus, alternative delivery means consist of the administration of genetic material to instruct the target cells to produce the editing machinery. Transfection of plasmids encoding the base editor and the gRNA is simple and inexpensive. However, this technique has the disadvantage of being toxic for the cells[125]. On the contrary, mRNA transfection allows a rapid and more transient expression of the enzyme and has been shown to result in high base editing efficiencies in HSPCs[126–129] (**Annex 1**).

a) Strategies aiming to correct the SCD mutation

(1) CRISPR/Cas9-based approaches

Due to its high prevalence and its unique causal mutation, SCD is a good target for the development of gene correction gene therapy strategies. Efficient *HBB* gene correction has been demonstrated by HDR using CRISPR/Cas9[130–132] (**Figure 21**). However, homologous recombination events compete with other repair mechanisms such as NHEJ (see **I.E.3.a) Editing byproducts at the on-target site**). Thus, in these studies, correction of the mutation is accompanied by a significant rate of alleles harboring InDels. This observation raises the risk of inducing a β-thalassemic phenotype by frameshift mutations that abolish the β-globin expression.

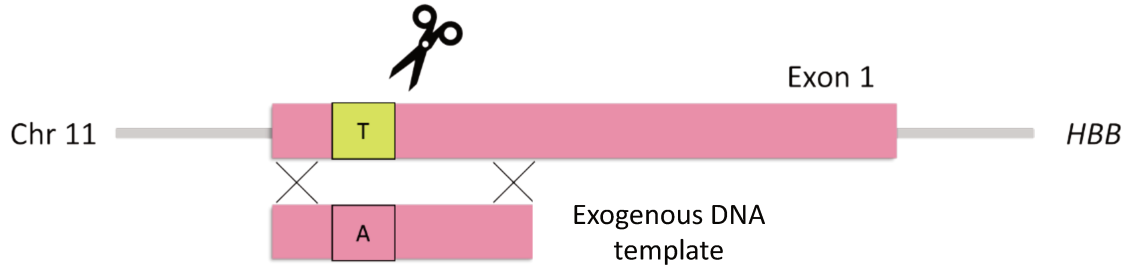


Figure 21. Correction of the SCD mutation by homologous recombination.

The Cas9 nuclease, delivered with an exogenous DNA template homologous to the cleavage site, allows the insertion of specific mutations into the targeted sequence. Thus, the mutation responsible for SCD in the β-globin gene can be corrected by HDR with a donor DNA template carrying the wild-type allele.

In 2021, 2 phase 1/2 clinical trials aiming to demonstrate the safety of a gene therapy strategy to correct the SCD mutation in patients HSCs have been announced. The first, a clinical trial conducted by the University of California, will include 9 patients (NCT04774536)[133,134]. The second, conducted by the company Graphite Bio, will include approximately 15 patients (NCT04819841)[132,135].

(2) Base editing-based approaches

The A>T substitution responsible for SCD leads to the replacement of a glutamic acid (codon GAG) by a valine (codon GTG) in the 6th position of the β-globin chain. As ABEs can convert A●T base pairs to G●C, they only allow the valine GTG codon to be converted to an alanine-encoding GCG codon. Although this conversion does not restore the original amino acid, it does introduce an amino acid found in a non-pathogenic variant of HbA: HbG-Makassar (**Figure 22**)[136–139]. While the SCD mutation was difficult to reach with the first generation of base editors, in 2019 Miller et al. developed a base editor that recognizes alternative PAM sequences allowing for the first time to create the HbG-Makassar mutation[80] (see **I.F.2 Genome editing with base editors: requirements and optimizations**). In 2021, Newby et al. increased the efficiency of this enzyme and achieved 80% mutation correction in SCD HSPCs[127]. They demonstrated that this strategy allows the correction of the

SCD phenotype *in vitro*, but also *in vivo* after autologous transplantation of edited HSCs in a SCD mouse model[127].

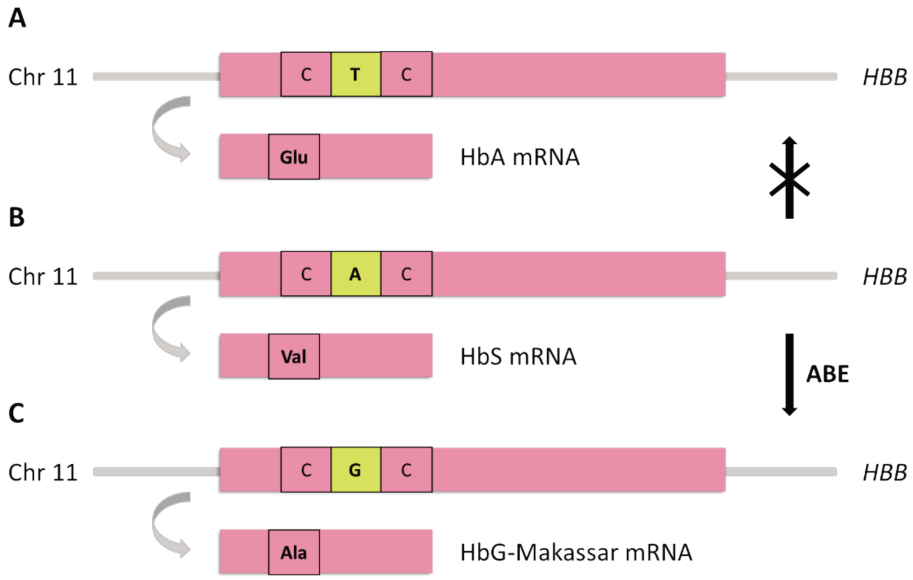


Figure 22. Correction of the SCD mutation by an ABE (A>G) results in the production of the HbG-Makassar variant.

(**A**) The wild-type β-globin gene encodes for a glutamic acid at the 6th position of the protein. (**B**) The mutated β^S-globin gene encodes for a valine at the 6th position of the protein. (**C**) The β^G-globin variant gene encodes for an alanine at the 6th position of the protein.

b) Strategies aiming to reactivate HbF expression

SCD and β-thalassemia patients with HPFH syndrome have an attenuated phenotype due to the protective effect of HbF (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**). Thus, several strategies aim at reactivating HbF expression, either by decreasing the expression of transcription factors repressing *HBG1/2* genes (e.g., BCL11A), or by disrupting the binding sites of these factors at *HBG1/2* promoters (e.g., binding sites of BCL11A and LRF) (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**). Similarly, other strategies aim to recruit transcription activators (e.g., KLF1 and TAL1) to *HBG1/2* promoters (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**)[140](**Annex 1**).

(1) Targeting the *BCL11A* erythroid enhancer

(a) CRISPR/Cas9-based approaches

BCL11A is a transcription factor that strongly represses HbF expression (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**)[141,142]. A therapeutic strategy aiming to fully abolish the expression of this factor would be toxic because it is essential for HSC function and lymphocyte development[143]. Thus, the erythroid-specific enhancer region of *BCL11A* has been shown to be an attractive therapeutic target[144]. This region, located 58 kb downstream of the TSS in the second intron of *BCL11A*, has a binding site for the transcriptional activator GATA1. The Cas9 nuclease can be used to disrupt this binding site by introducing InDels. This strategy specifically represses BCL11A expression in the erythroid lineage and thus decreases the inhibition that this factor has on HbF expression (**Figure** 23). Preclinical testing in human HSPCs and by autologous transplant in rhesus monkey showed very high levels of InDel maintained over time in parallel with therapeutic HbF re-expression[144,145].

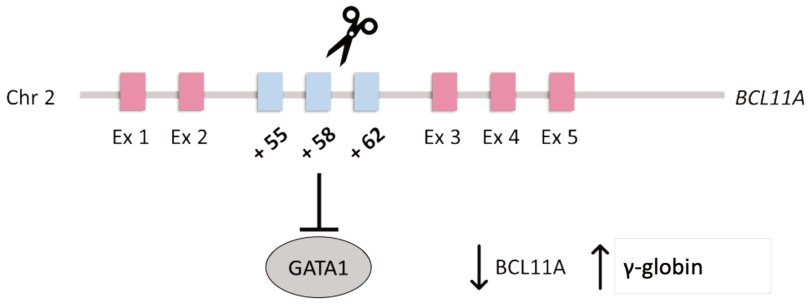


Figure 23. CRISPR/Cas9 nuclease InDels induction at the erythroid-specific enhancer of *BCL11A* to derepress HbF.

Disruption of the GATA1 binding site by Cas9 nuclease in the *BCL11A* erythroid-specific enhancer decreases its expression and thus derepresses HbF production.

A similar strategy uses ZFN to achieve the disruption of the *BCL11A* erythroid enhancer[146,147]. The first clinical trial based on this strategy rely on ZFN and was launched by Sangamo Therapeutics in patients with SCD and transfusion-dependent β-thalassemia (NCT03653247 and NCT03432364, respectively). While results are promising for the treatment of SCD, HbF levels obtained after treatment were not sufficient and patients resumed RBC transfusions leading to the end of recruitment for this trial[148,149].

The preclinical results based on the equivalent CRISPR strategy paved the way for two phase 1/2 clinical trials sponsored by Vertex Pharmaceuticals (NCT03655678 and NCT03745287) for the treatment of transfusion-dependent β-thalassemia and SCD by targeting the *BCL11A* enhancer. In 2021, Frangoul et al. published the first results of these trials in a β-thalassemia patient and a SCD patient[112]. More than one year after treatment, both patients showed a high level of editing in bone marrow cells and peripheral blood nucleated cells, confirming long-term graft maintenance. The thalassemic patient showed anemia correction, allowing the cessation of transfusions 1 month after treatment. The SCD patient did not experience any further vaso-occlusive crises during the follow-up period and was able to discontinue transfusions a few days after treatment. Interestingly, both patients showed pancellular expression of HbF after treatment. These results allowed the initiation of a phase 3 clinical trial in a dozen of pediatric patients (NCT05329649). Finally, an observational study including patients from these 3 studies will evaluate efficacy and toxicity in the 15 years following treatment (NCT04208529).

Of note, Novartis launched a clinical trial using a similar strategy (reduction of *BCL11A* erythroid expression) with no results reported to date (NCT04443907)[150].

(b) Base editing-based approaches

As demonstrated with CRISPR/Cas9 studies, disruption of the transcription factor GATA1 binding site in the erythroid enhancer of *BCL11A* is an efficient strategy to induce HbF re-expression (see **I.G.3.b)(1)(a) CRISPR/Cas9-based approaches**). Similarly, Zeng et al. used a CBE to introduce a C>T conversion within this motif (**Figure 24**)[126].

This technique allows editing of the targeted cytosine with very high efficiencies in HSPCs of β-thalassemia or SCD patients[126]. *In vitro* assessment of this strategy showed substantial HbF reactivation and subsequent correction of the phenotype. Moreover, *in vivo* experiments showed high maintenance of the edits over time[126].

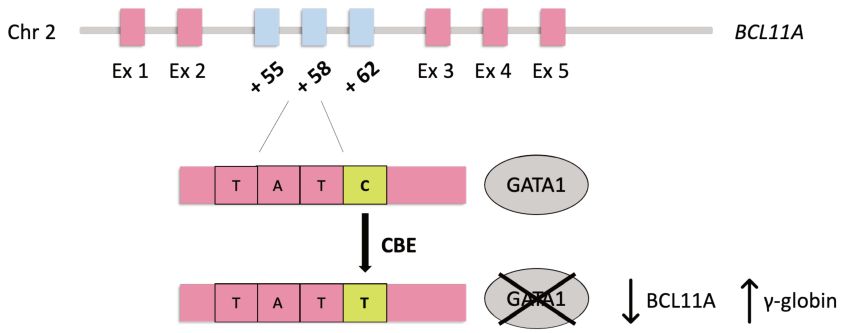


Figure 24. CBE usage allows the disruption of the GATA1 binding site in the *BCL11A* erythroid enhancer.

CBE can be used to disrupt the GATA1 transcriptional activator binding site in the *BCL11A* erythroid enhancer. This strategy allows HbF reactivation by decreasing the expression of its transcriptional repressor BCL11A.

(2) Targeting regulators binding sites in the *HBG* genes

(a) CRISPR/Cas9-based approaches

Several HPFH mutations found in the -115 region of the *HBG* promoters disrupt the binding site of the potent HbF repressor BCL11A[9] (**Figure 4**). Several publications show the efficiency of Cas9 nuclease to induce InDels in this region, reactivate HbF and correct *in vitro* the SCD phenotype[151,152] (**Figure 25A**). *In vivo* mice experiments demonstrated the maintenance of high InDel rates at this locus over time[151]. In 2019, Humbert et al. transplanted autologous edited HSPCs in rhesus macaques[153]. They demonstrated the long-term maintenance of edited cells and persistent HbF reactivation in peripheral RBCs after one year[153].

This strategy has also been brought to clinical evaluation for SCD patients by Editas Medicine, with no published results to date[154,155] (NCT04853576).

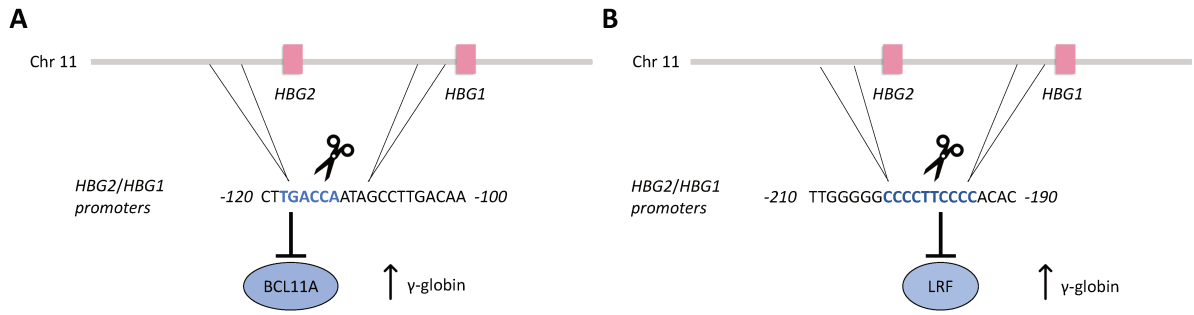


Figure 25. CRISPR/Cas9 induction of InDels in *HBG1/2* promoters.

CRISPR/Cas9 can be used to disrupt HbF repressors binding sites in order to reactivate HbF expression. Here, InDels are generated in the -115 region (**A**) or in the -200 region (**B**) of *HBG* promoters, in order to disrupt the BCL11A or the LRF repressor binding site, respectively.

Similarly, certain HPFH mutations located in the -200 region of *HBG1/2* promoters have been described to disrupt the binding site of LRF, a potent repressor of HbF expression[9] (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**). While the abolition of LRF expression allows re-expression of HbF, it has not been considered as a therapeutic strategy because it delays erythroid differentiation[10]. In 2020, we reproduced the effect of HPFH mutations in *HBG* promoters by inducing InDels in the -200 region with CRISPR/Cas9[156] (**Figure 25B**). Applied in SCD HSPCs, this strategy demonstrated its ability to significantly reactivate HbF and maintain high levels of InDel over time after *in vivo* transplantation. Interestingly, it also showed efficacy for correction of the β-thalassemic phenotype in cells from patients with the most severe β^0/β^0 genotype[157].
However, simultaneous Cas9 nuclease cutting of both *HBG* promoters can generate deletion of the intergenic 4.9-kb region, with uncertain consequences[151,158].

(b) Base editing-based approaches

Regarding strategies aiming to mimic HPFH mutations, it is interesting to note that the use of base editors not only allows the disruption of putative repressors binding sites -as by using Cas9 nuclease- but also the creation of activator binding sites[129,140] (**Annex 1**). In fact, these tools have demonstrated their efficacy to modify the bases composing BCL11A and LRF repressors binding sites but also to create binding sites

for the TAL1, GATA1, and KLF1 activators[82,83,159–162] (see **I.A.3 Hereditary Persistence of Fetal Hemoglobin**). Some of these strategies have been shown to be effective in HSPCs and have led to the re-expression of HbF in their erythroid progeny [159,160]. Mayuranathan et al. demonstrated correction of the SCD phenotype by creating the TAL1 binding site[140]. Similarly, we demonstrated the efficacy of the KLF1 binding site creation at the -200 locus and consequent correction of the SCD and β-thalassemic phenotypes (**Annex 1**). Finally, a recent publication highlighted the potential of simultaneous T>C base editing in positions -123/-124 of *HBG* promoters to increase HbF potentially by recruiting the KLF1 transcription activator[129].

II. Aim of the project

β-hemoglobinopathies are the most common inherited monogenic disorders in the world. Allogenic bone marrow transplant is the only curative treatment but suffers from the poor availability of compatible donors. Transplantation of autologous, genetically modified HSPCs was investigated as a treatment option for β-thalassemia and SCD patients. However, current gene therapy strategies based on the use of lentiviral vectors or CRISPR/Cas9 nuclease are not equally effective in all the patients and/or raise safety concerns. In particular, vector integration can cause insertional mutagenesis and CRISPR/Cas9 mediated DSB generation can induce cell toxicity and genomic rearrangements. Base editors are CRISPR/Cas9-derived tools allowing precise mutations insertion without DSB generation.

The aim of this thesis is to use base editing to develop safer strategies for the treatment of β-hemoglobinopathies. Therefore, we designed DSB-free base editing strategies to precisely (i) correct three prevalent and severe β-thalassemia mutations; (ii) correct the SCD mutation.

(i) Correction of three prevalent and severe β-thalassemia mutations.

Despite the large variety of β-thalassemia mutations, they are very unequally distributed as only 20 prevalent alleles account for more than 80% of the mutations worldwide[163]. Thus, we focused our attention on three prevalent and severe β-thalassemia G>A point mutations that could theoretically be corrected using ABEs.

We aimed to identify an efficient base editing strategy to correct each selected mutation. With this aim, we designed different combinations of gRNAs and ABEs and screened them in T cells obtained from β-thalassemia patients harboring the selected mutations. For each mutation, we selected the best performing combination, which was then used in patients' HSPCs, the target cell population for gene therapy of β-hemoglobinopathies.

Evaluation of the therapeutic potential of these strategies was performed in RBCs derived from edited β-thalassemic HSPCs. The base editing machinery was delivered to HSPCs using a clinically relevant delivery system, i.e. RNA transfection.

Cells were then *in vitro* differentiated toward the erythroid lineage in order to evaluate the correction of the β-thalassemic phenotype in RBCs.

In parallel, we aimed at assessing the safety of these strategies by evaluating the effect of bystander edits and off-target activity caused by the selected gRNA/ABE combinations.

(ii) Correction of the SCD mutation.

SCD is caused by a T>A mutation, which cannot be precisely corrected with current base editing tools. However, A>G conversion of the mutation with ABE allows the replacement of the valine in position 6 (the amino acid responsible for Hb polymerization) by an alanine. This leads to the production of a known Hb variant called HbG-Makassar, which was described in homozygous healthy individuals.

Thus, we aimed to develop an efficient way to induce the HbG-Makassar mutation. As the SCD mutation is not easily accessible due to target sequence restrictions of the first-generation ABEs, we aimed to develop a strategy to efficiently target this mutation by sequentially optimizing ABE constructs and gRNAs.

III. Adenine base editor-mediated correction of a common and severe β-thalassemia mutation

Adenine base editor-mediated correction of a common and severe β-thalassemia mutation

Giulia Hardouin[1,2,3], Panagiotis Antoniou[1], Pierre Martinucci[1], Sandra Manceau[4], Laure Joseph[3,4], Cécile Masson[5], Marina Cavazzana[2,3,4], Annarita Miccio[1*]

[1]Laboratory of chromatin and gene regulation during development, Imagine Institute, INSERM UMR1163, Université Paris Cité, Paris, France
[2]Laboratory of Human Lymphohematopoiesis, Imagine Institute, INSERM UMR 1163, Université Paris Cité, Paris, France
[3]Biotherapy Clinical Investigation Center, Necker Children's Hospital, Assistance Publique Hopitaux de Paris, Paris, France
[4]Biotherapy Department, Necker Children's Hospital, Assistance Publique Hopitaux de Paris, Paris, France
[5]Bioinformatics Platform, Imagine Institute, INSERM UMR1163, Paris Cité University, Paris, France.

*To whom correspondence should be addressed: A.M., Imagine Institute, 24, Boulevard du Montparnasse, 75015 Paris, France. E-mail address: annarita.miccio@institutimagine.org. Phone number: +33 1 42 75 43 34

A. Abstract

β-thalassemia is one of the most common genetic diseases worldwide and is caused by mutations affecting β-globin production. The only curative treatment is allogenic hematopoietic stem/progenitor cells (HSPCs) transplantation, an approach limited by compatible donor availability and immunological complications. Therefore, transplantation of autologous, genetically modified HSPCs is an attractive therapeutic option. However, current gene therapy strategies based on the use of lentiviral vectors or CRISPR/Cas9 nuclease are not equally effective in all the patients and/or raise safety concerns. Thus, base editing strategies aiming to correct the genetic defect in patients HSPCs could provide a safe and effective treatment. Here, we developed a strategy to correct one of the most prevalent β-thalassemic mutations [IVS1-110 (G>A)] using the SpRY-ABE8e base editor. RNA delivery of the base editing system was safe and led to ~80% of gene correction in β-thalassemic patients' HSPCs without causing dangerous double-strand DNA breaks. In HSPC-derived erythroid populations, this strategy was able to restore β-globin production and correct inefficient erythropoiesis and apoptosis typically observed in β-thalassemia. In conclusion, this proof-of-concept study paves the way for the

development of a safe and effective autologous gene therapy approach for β-thalassemia.

B. Introduction

β-thalassemia is a highly prevalent[164] recessive disorder caused by mutations in the β-globin gene (*HBB*) locus that reduce (β^+) or eliminate (β^0) β-globin production. Depending on the genotype, the clinical phenotype can vary from mild to transfusion-dependent anemia.

IVS1-110 (G>A) is one of the most common mutations in the Middle East and Mediterranean area[165–167]. This mutation leads to the formation of a splice acceptor site in the first *HBB* intron causing partial intron retention and generation of a premature termination codon in 90% of the β-globin transcripts[168,169]. Due to the low levels of correctly spliced mRNAs, and the inhibition of their translation caused by the aberrant transcripts, this mutation leads to a strong β-globin downregulation[169]. Thus, IVS1-110 is a severe β^+ mutation, and homozygous patients or compound heterozygotes harboring this mutation and a β^0 mutation have a clinical phenotype similar to β^0/β^0 patients[118].

Transplantation of autologous, genetically modified HSPCs was investigated as a treatment option for β-thalassemia patients[2]. Lentiviral-mediated gene addition is modestly effective in patients harboring the IVS1-110 mutation[118].

Editing approaches use nucleases, such as CRISPR/Cas9 that induces DNA double-strand DNA breaks (DSBs) via a guide RNA (gRNA) complementary to the genomic target. The DSB can be repaired via homologous-directed repair (HDR) by providing a donor template containing the wild type sequence, allowing gene correction. However, HDR is poorly efficient in hematopoietic stem cells[66]. Nuclease-based strategies based on the disruption of the splice site generated by the IVS1-110 mutation through InDel formation after non-homologous end-joining-mediated DSB repair are effective in restoring correct hemoglobin expression[170,171]. However, only a fraction of InDels abolish the aberrant splice site[171]. Furthermore, DSBs can induce cell toxicity[172] and genomic rearrangements[61].

Base editing allows precise DNA repair without generating DSBs. Adenine base editors (ABE) contain a Cas9 nickase and a deaminase, allowing A>G conversions[75].

Here, we exploited ABEs to correct the IVS1-110 mutation in patients' HSPCs and the pathological phenotype of HSPC-derived erythroid cells.

C. Materials and Methods

Briefly, we purified $CD34^+$ and T cells from β-thalassemia patients' blood. $CD34^+$ HSPCs were differentiated towards the erythroid lineage[156]. Globin and hemoglobin expression was evaluated by RT-qPCR and HPLC[156]. Correction of the phenotype was assessed by flow cytometry and by quantitative phase microscopy single-cell analysis[173]. The on-target and *in silico* predicted off-targets were subjected to Sanger sequencing and/or NGS. Detailed protocols are described in **Supplemental Methods**.

D. Results and discussion

We designed 6 gRNAs (**Figure 1A, Supplemental Table 1**) compatible with ABE8e and NG-ABE8e[82] (recognizing NGG and NG PAM, respectively), and SpRY-ABE8e[174] that we generated by combining the highly processive deaminase from ABE8e[82] with the SpRY PAM-less Cas9 nickase[81]. We screened gRNA/BE combinations in T cells obtained from β-thalassemia patients homozygous for the IVS1-110 mutation (BT0 and BT1, **Figure 1B**). gRNA1/SpRY-ABE8e was the most efficient combination able to correct the mutation in nearly 90% of the *HBB* mutant alleles, as evaluated by Sanger sequencing (**Figure 1C**). Two bystander edits were observed at positions 6 and 8, including a C>G conversion occurring at relatively low frequency (**Figure 1D**), as previously reported at other loci with ABEs[87,175]. Importantly, these changes have likely no consequences on *HBB* expression as they occur in an intronic region and SNPs described at these positions are not associated with any clinical manifestations.

gRNA1 and SpRY-ABE8e mRNA were transfected in HSPCs from two homozygous patients (BT0 and BT1) and one compound heterozygous harboring the IVS1-110 and a β^0 mutation (BT2) (**Figure 1B-1E**). Gene correction was high and similar between the three donors with little InDel generation as evaluated by Sanger sequencing and NGS (**Figure 1F, Supplemental Figure 1A-C**). The high percentage

of bystander edit at position 8 in the heterozygous donor indicates that the gRNA also targets the allele containing the β^0 mutation, likely because the mismatch is located at the 5' end of the gRNA, where it is more easily tolerated[65] (**Supplemental Figure 1A**). NGS analysis confirmed a good product purity with most of the events being A>G conversions and a lower editing of cytosines leading mainly to C>G, and more rarely to C>T conversions (**Supplemental Figure 1D)**.

To evaluate the safety profile of our strategy, we performed NGS of the top-10 *in silico* predicted off-targets (**Supplemental Table 2**). Base editing was observed only at one site that, however, mapped to a region devoid of known genes and did not impact cell viability and differentiation (**Supplemental Figure 1B**). Off-target activity could be potentially mitigated using high-fidelity base editors or modulating DNA exposure time to the editing machinery. No InDels were detected at off-target sites, thus minimizing the possibility of DSB-induced genomic rearrangements (e.g., translocations).

We observed a similar number of progenitors before and after the treatment with no difference in the proportion of erythroid (BFU-E) and granulomonocytic (CFU-GM) colonies, demonstrating the safety of our strategy **(Figure 1E, Supplemental Figure 2A)**. Importantly, all the progenitors were edited with low to absent InDels (**Supplemental Figure 2B)**.

Transfected β-thalassemic and control healthy donor (HD) HSPCs were differentiated toward the erythroid lineage (**Figure 1E**). RT-PCR showed restoration of the correct β-globin splicing and a substantial decrease of aberrant transcripts characterized by lower mobility due to partial intron retention (**Figure 1G-H**). Wild-type β-globin mRNA levels were similar to those observed in HD cells for homozygous donors, while representing ~50% of the HD β-globin transcripts for the compound heterozygote (**Figure 1H**). In untreated β-thalassemic red blood cells (RBCs), HPLC showed elevated α-/non-α-globin ratios and low β-globin expression poorly compensated by fetal γ (γ^A+γ^G)-globins (**Figure 1I**). After treatment, RBCs exhibited higher levels of β-globin chain and HbA (**Figure 1I-J**). In edited cells from homozygous patients, the α-/non-α-globin ratio was comparable to the values measured in HD cells and substantially ameliorated in the erythroid cells from the compound heterozygote

(**Figure 2I**). To evaluate if bystander edits have an impact on *HBB* expression, we induced bystander edits in HD HSPCs achieving frequencies similar to those detected in the β-thalassemic samples and differentiated them towards the erythroid lineage (**Supplemental Figure 3A**). We observed no decrease of *HBB* expression neither at mRNA level nor at the protein level with a normal α-/non-α globin ratios (**Supplemental Figure 3B-C**).

In β-thalassemia, the imbalance between α- and β-globin production leads to precipitation of α-globins causing apoptosis of erythroid precursors and ineffective erythropoiesis[176]. Flow cytometry analysis of CD36 and CD71 early erythroid markers showed a proper downregulation in treated samples, demonstrating that the typical delayed erythroid differentiation of β-thalassemic cells was efficiently corrected (**Figure 2A-C**). The production of Reactive Oxygen Species (ROS) observed in β-thalassemic control erythroid cells was significantly decreased in treated samples **(Figure 2D**). Importantly, flow cytometry showed rescue of edited β-thalassemic erythroid precursors from apoptosis (**Figure 2E**).

In treated samples, we observed an increased enucleation rate along the differentiation, reaching levels similar to those observed in HD at the end of the differentiation (**Figure 2F**). Flow cytometry showed that the size of enucleated cells (typically reduced in β-thalassemic cell cultures, namely microcytosis) was increased in treated samples (**Figure 2G**). Finally, we used phase microscopy to evaluate several physical parameters in a quantitative manner (**Figure 2H, Supplemental Figure 4A-E**). β-thalassemic RBCs were characterized by reduced perimeter, surface and volume due to the microcytosis, and decreased dry mass and surface density due to the low Hb content. These parameters were normalized or substantially improved in treated RBCs.

Overall, our study demonstrates the efficacy of a base editing approach to treat patients with prevalent and severe β-thalassemic mutations (see also **Supplemental Discussion**). Validation of these results in HSPCs *in vivo* will provide sufficient proof of efficacy and safety to enable the clinical development of base-edited HSPCs for gene therapy of β-thalassemia.

E. Figures

Figure 1

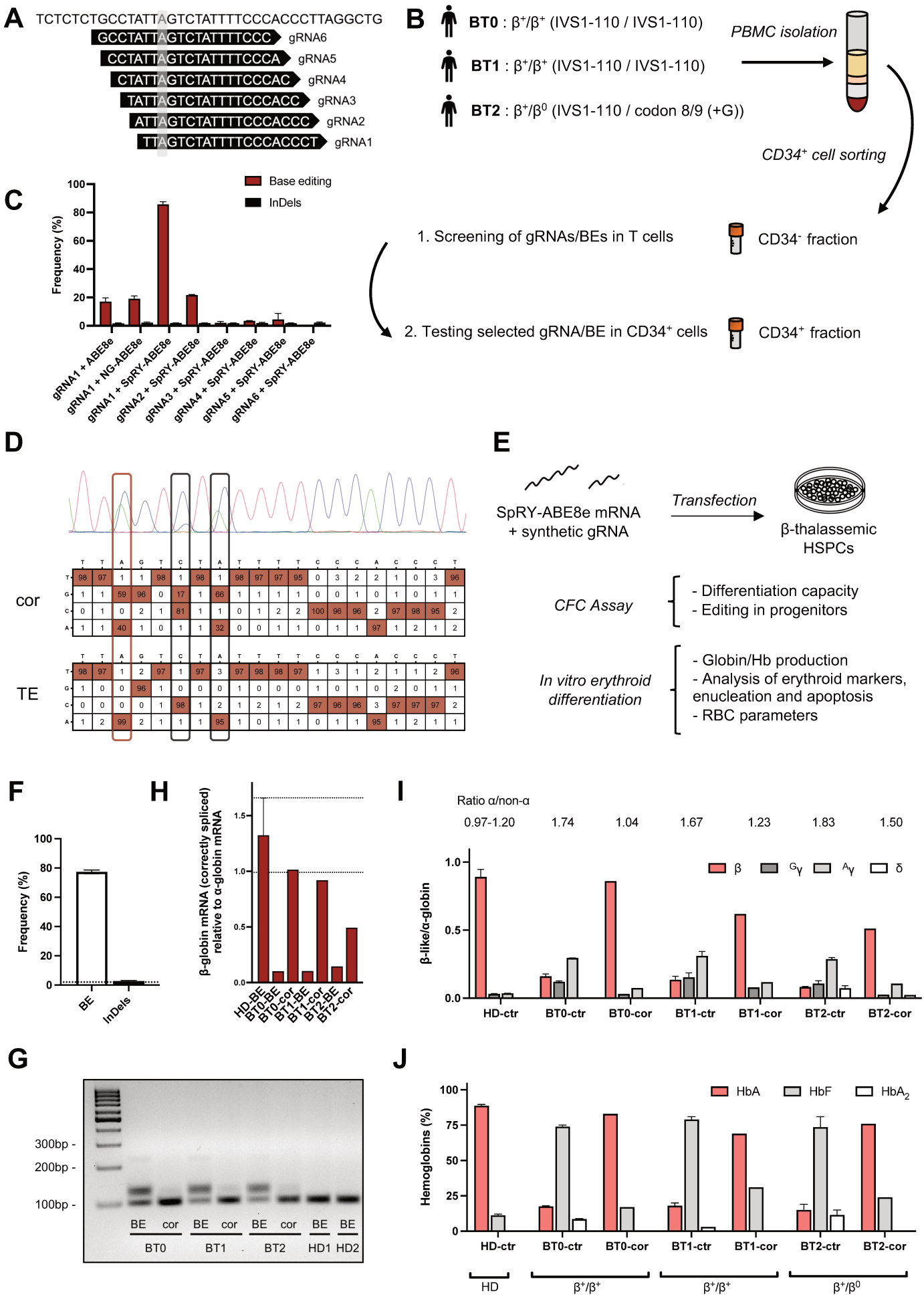
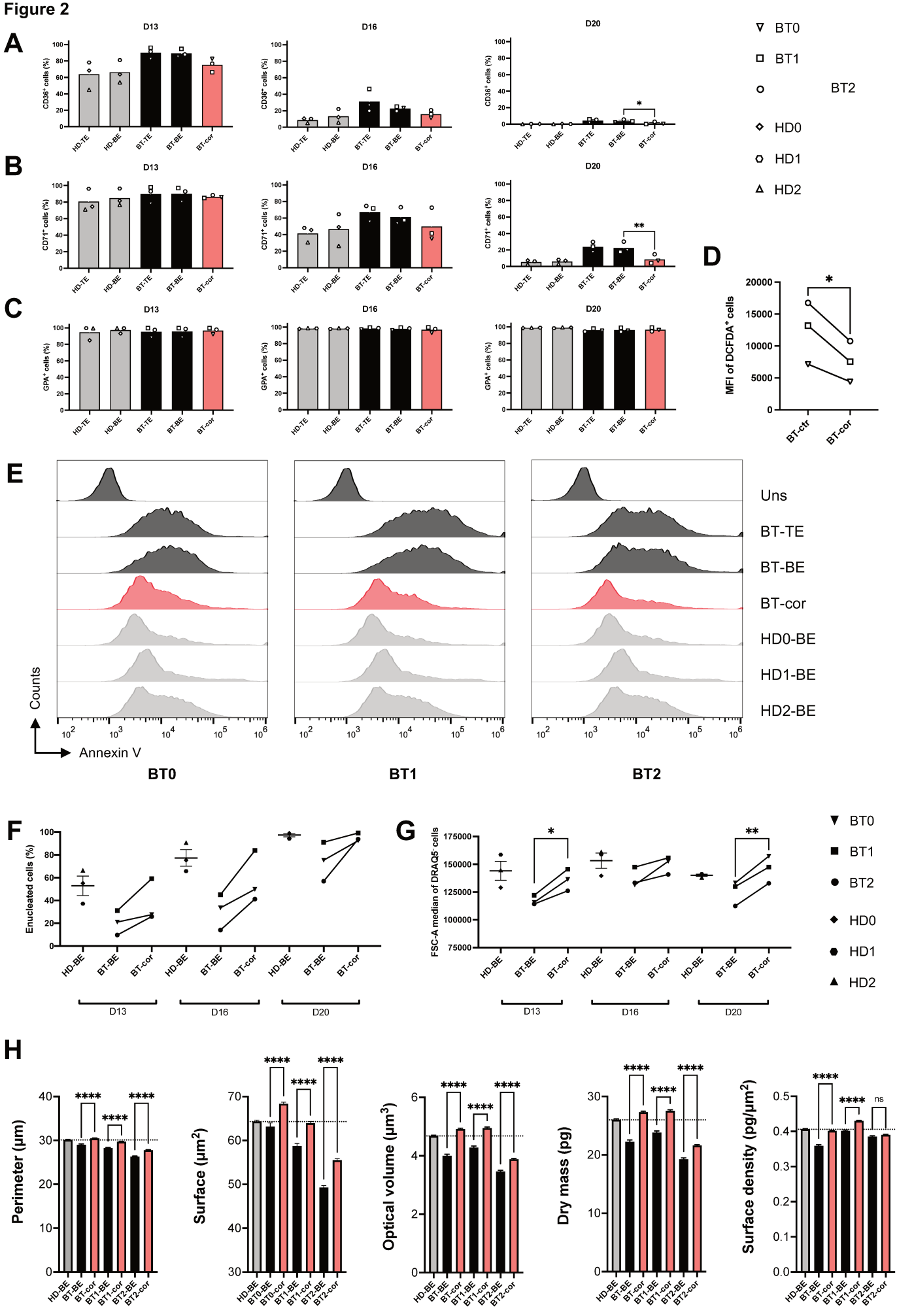

A
TCTCTCTGCCTATTAGTCTATTTTCCCACCCTTAGGCTG
GCCTATTAGTCTATTTTCCC gRNA6
CCTATTAGTCTATTTTCCCA gRNA5
CTATTAGTCTATTTTCCCAC gRNA4
TATTAGTCTATTTTCCCACC gRNA3
ATTAGTCTATTTTCCCACCC gRNA2
TTAGTCTATTTTCCCACCCT gRNA1
B
BT0 : β+/β+ (IVS1-110 / IVS1-110)
BT1 : β+/β+ (IVS1-110 / IVS1-110)
BT2 : β+/β0 (IVS1-110 / codon 8/9 (+G))
PBMC isolation
CD34+ cell sorting
C
Base editing
InDels
Frequency (%)
gRNA1 + ABE8e
gRNA1 + NG-ABE8e
gRNA1 + SpRY-ABE8e
gRNA2 + SpRY-ABE8e
gRNA3 + SpRY-ABE8e
gRNA4 + SpRY-ABE8e
gRNA5 + SpRY-ABE8e
gRNA6 + SpRY-ABE8e
1. Screening of gRNAs/BEs in T cells
CD34- fraction
2. Testing selected gRNA/BE in CD34+ cells
CD34+ fraction
D
cor
TE
E
SpRY-ABE8e mRNA
+ synthetic gRNA
Transfection
β-thalassemic
HSPCs
CFC Assay
- Differentiation capacity
- Editing in progenitors
In vitro erythroid differentiation
- Globin/Hb production
- Analysis of erythroid markers, enucleation and apoptosis
- RBC parameters
F
Frequency (%)
BE
InDels
H
β-globin mRNA (correctly spliced) relative to α-globin mRNA
HD-BE
BT0-BE
BT0-cor
BT1-BE
BT1-cor
BT2-BE
BT2-cor
I
Ratio α/non-α
0.97-1.20
1.74
1.04
1.67
1.23
1.83
1.50
β-like/α-globin
β
Gγ
Aγ
δ
HD-ctr
BT0-ctr
BT0-cor
BT1-ctr
BT1-cor
BT2-ctr
BT2-cor
G
300bp -
200bp -
100bp -
BE cor BE cor BE cor BE BE
BT0 BT1 BT2 HD1 HD2
J
Hemoglobins (%)
HbA
HbF
HbA2
HD-ctr
BT0-ctr
BT0-cor
BT1-ctr
BT1-cor
BT2-ctr
BT2-cor
HD
β+/β+
β+/β+
β+/β0

Figure 2


A
B
C
D
E
F
G
H
D13
D16
D20
CD36+ cells (%)
CD71+ cells (%)
GPA+ cells (%)
HD-TE
HD-BE
BT-TE
BT-BE
BT-cor
BT0
BT1
BT2
HD0
HD1
HD2
MFI of DCFDA+ cells
BT-ctr
Uns
BT-TE
BT-BE
BT-cor
HD0-BE
HD1-BE
HD2-BE
Counts
Annexin V
Enucleated cells (%)
FSC-A median of DRAQ5+ cells
Perimeter (µm)
Surface (µm2)
Optical volume (µm3)
Dry mass (pg)
Surface density (pg/µm2)
ns

F. Figure legends

Figure 1. A base editing strategy to efficiently correct the IVS1-110 (G>A) mutation in β-thalassemic HSPCs and restore normal Hb production.

A. gRNAs 1-6 were manually designed to place the IVS1-110 (G>A) mutation in position 3 to 8 of the protospacer. The mutation is highlighted with a grey box. **B.** Overview of the cell collection for testing the ability of gRNA/BE combinations to correct the IVS1-110 (G>A) mutation. Peripheral blood mononuclear cells (PBMCs) were isolated from 2 patients homozygous for the IVS1-110 (G>A) mutation and 1 compound heterozygous patient harboring the IVS1-110 (G>A) mutation and the codon 8/9 (+G) β^0 mutation. After $CD34^+$ cell sorting, T cells were recovered from the negative fraction for testing gRNA/BE combinations, before moving to $CD34^+$ cells with the best-performing strategy. **C.** Frequency of corrected alleles (normalized to the frequency of GFP^+ cells) as evaluated by EditR and InDels frequency as assessed by TIDE in T cells transfected with different combinations of synthetic gRNAs and ABE mRNAs. Data are expressed as mean±standard error of the mean (SEM) (n=3, 2 donors). **D.** Representative percent composition of Sanger sequencing traces measured to be significantly different from noise (in red), as assessed by EditR following Sanger sequencing in T cells transfected with gRNA1/SpRY-ABE8e (cor) or with TE buffer (TE). Target base position is highlighted with a red box and the observed bystander edits with black boxes. SNPs rs777028217 (G>A) and rs1480884739 (T>C) map to positions 6 and 8 are not associated with a clinical phenotype. **E.** Experimental protocol used for base editing experiments in β-thalassemic HSPCs. SpRY-ABE8e mRNA and synthetic gRNA1 were co-transfected in β-thalassemic HSPCs. Cells were plated in a methylcellulose-based medium under conditions supporting erythroid (BFU-E) and granulomonocytic (CFU-GM) differentiation, or differentiated into mature RBCs using a three-phase erythroid differentiation protocol. **F.** Frequency of corrected alleles (BE) as evaluated by EditR and InDels frequency as assessed by TIDE in β-thalassemic HSPCs. Data are expressed as mean±SEM (3 donors). Frequency of corrected alleles in the cells obtained from the compound heterozygous patient (BT2) was corrected taking into account only the alleles harboring the IVS1-110 (G>A) mutation. The dotted line

represents the maximum background noise of InDels calculated by TIDE. **G.** RT-PCR using primers amplifying a region spanning the *HBB* exon 1-exon 2 junction. cDNA was obtained from erythroid cells derived from corrected β-thalassemic HSPCs (cor). As controls, we used erythroid cells derived from β-thalassemic or HD HSPCs transfected only with SpRY-ABE8e mRNA (BE) (3 β-thalassemia patients and 2 HDs). **H.** RT-qPCR using primers detecting exclusively correctly spliced β-globin mRNAs in erythroid cells derived from corrected β-thalassemic HSPCs (cor). As controls, we used erythroid cells derived from β-thalassemic or HD HSPCs transfected only with SpRY-ABE8e mRNA (BE) (3 β-thalassemia patients and 2 HD). β-globin expression was normalized to α-globin. Data are expressed as mean±SEM. Dotted lines indicate maximum and minimum values observed in HD cells. **I.** Expression of β-, ^{G}γ-, ^{A}γ- and δ- globin chains measured by RP-HPLC in β-thalassemic and HD RBCs. β-like-globin expression was normalized to α-globin. The α-/non-α-globin ratio is reported on top of the graph. RBCs were obtained from corrected β-thalassemic HSPCs (cor). As controls (ctr), we pooled data obtained in RBCs derived from β-thalassemic (BT) or HD HSPCs transfected with TE or with SpRY-ABE8e mRNA only (3 β-thalassemia patients and 2 HDs). Data are expressed as mean±SEM. **J.** Analysis of HbA, HbF and HbA_2 by CE-HPLC in β-thalassemic and HD RBCs. We calculated the percentage of each Hb type over the total Hb tetramers. RBCs were obtained from corrected β-thalassemic HSPCs (cor). As controls (ctr), we pooled data from RBCs derived from β-thalassemic or HD HSPCs transfected with TE or with SpRY-ABE8e mRNA only (3 β-thalassemia patients and 2 HDs). Data are expressed as mean±SEM. BT, β-thalassemia patients. HD, healthy donors.

Figure 2. **Efficient reversion of the IVS1-110 (G>A) mutation in β-thalassemic HSPCs corrects ineffective erythropoiesis.**

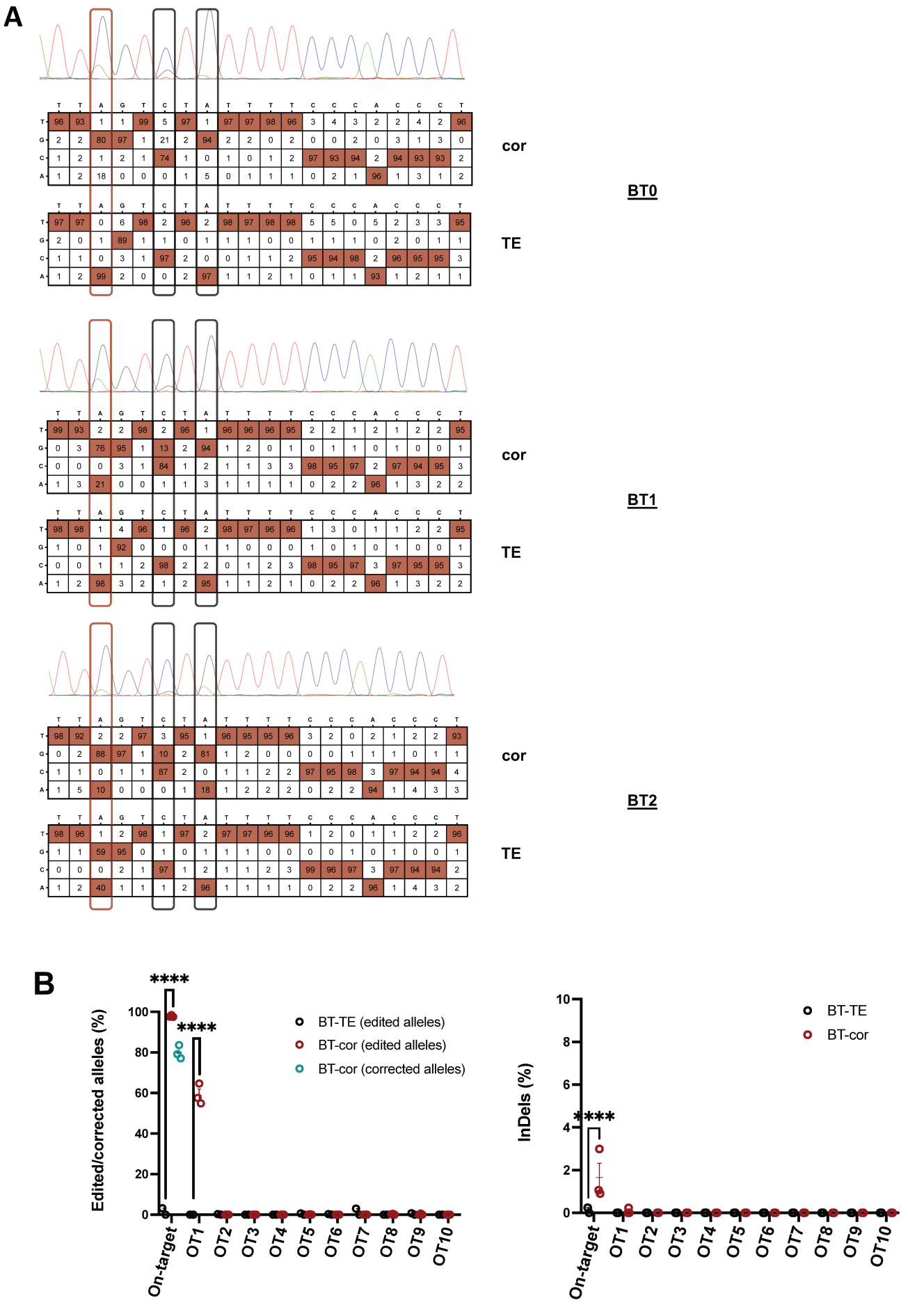
A-C. Frequency of $CD36^+$ (A), $CD71^+$ (B) and GPA^+ (C) cells at day 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis. Control BT or HD cells were either transfected with TE buffer (TE), or only with SpRY-ABE8e mRNA (BE). Data are expressed as mean±SEM (3 β-thalassemia patients and 3 HDs). * p≤0.05 ** p≤0.01 (Paired T-test, BT-BE *vs* BT-cor). **D.** Mean Fluorescence Intensity (MFI) in ROS-containing cells ($DCFDA^+$) for control (ctr) and edited (cor) β-thalassemic samples. For the controls (ctr), we pooled data obtained in RBCs derived

from β-thalassemic (BT) HSPCs transfected with TE or with SpRY-ABE8e mRNA only (3 β-thalassemia patients). * p≤0.05 (Paired T-test). **E.** Flow cytometry histograms showing the frequency of apoptotic cells (AnnexinV$^+$-cells) in the 7AAD$^-$ cell population in unstained (Uns), β-thalassemic and HD samples at day 13 of erythroid differentiation (3 β-thalassemia patients and 3 HDs). **F.** Frequency of enucleated cells at day 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of cells stained with the DRAQ5 nuclear dye (3 β-thalassemia patients and 3 HDs). Data for HD samples are expressed as mean±SEM. **G.** Cell size of enucleated cells at day 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of the median FSC-A intensity (3 β-thalassemia patients and 3 HDs). Data for HD samples are expressed as mean±SEM. * p≤0.05 ** p≤0.01 (Paired T-test, BT-BE *vs* BT-cor). **H.** Single RBC parameters (perimeter, surface, optical volume, dry mass and surface density) evaluated by quantitative phase image microscopy in RBCs obtained from corrected β-thalassemic HSPCs (cor). As controls, we used RBCs derived from β-thalassemic or HD HSPCs transfected only with SpRY-ABE8e mRNA (BE). Data are expressed as mean±SEM (3 β-thalassemia patients and 2 HDs). **** p≤0.0001 (Ordinary one-way ANOVA). D, day.

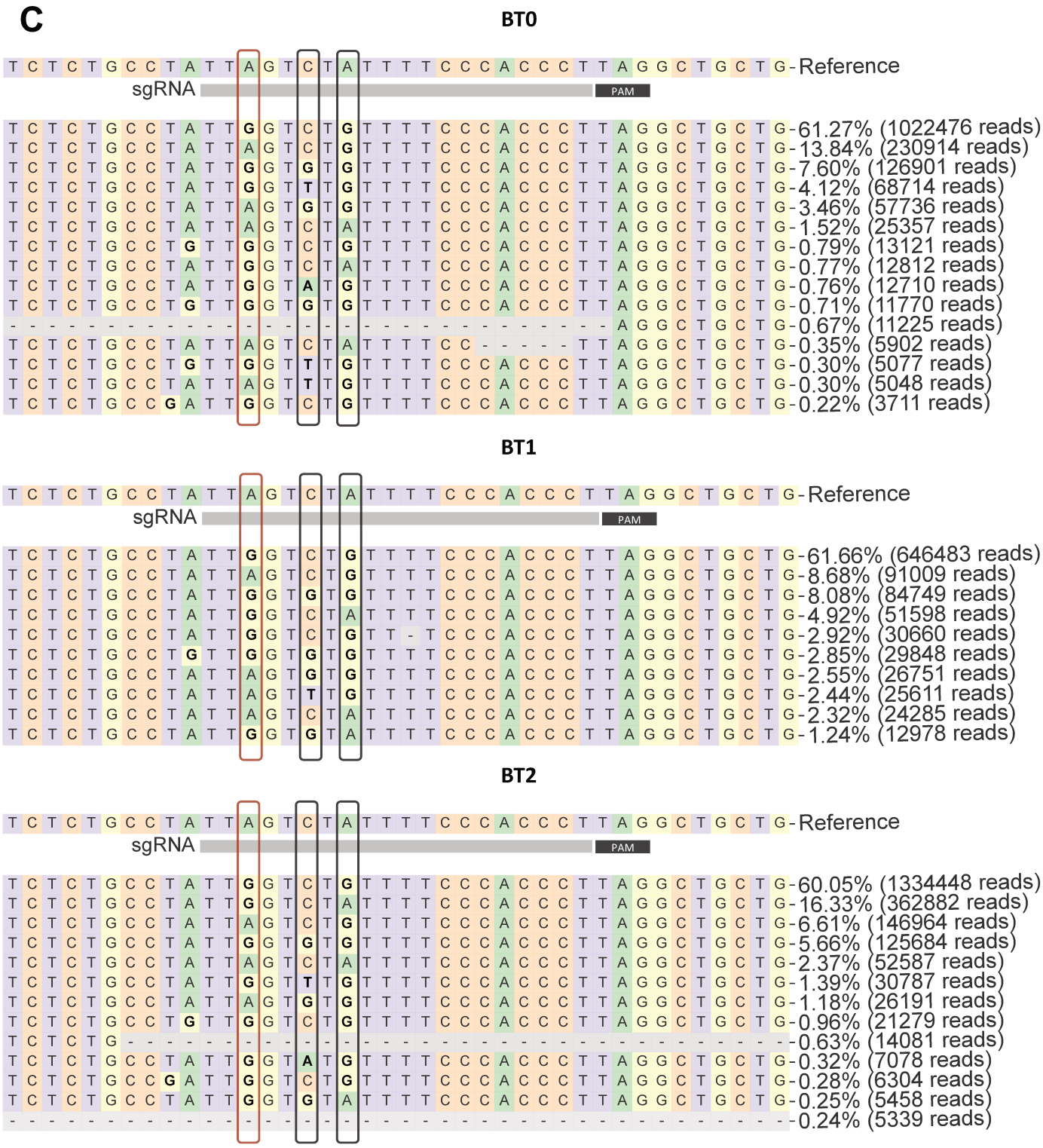
G. Supplemental files

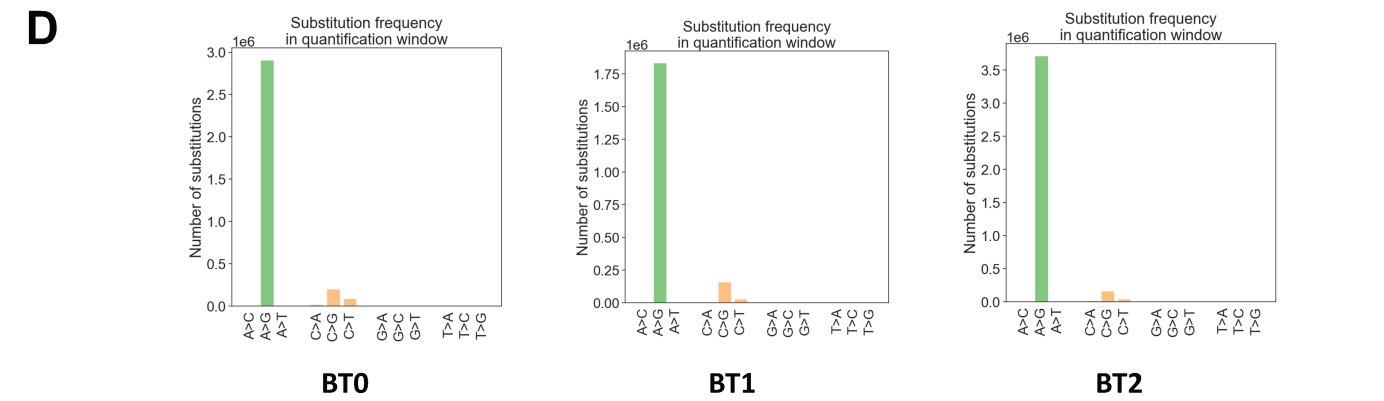
1. Supplemental figures

Supplemental Figure 1

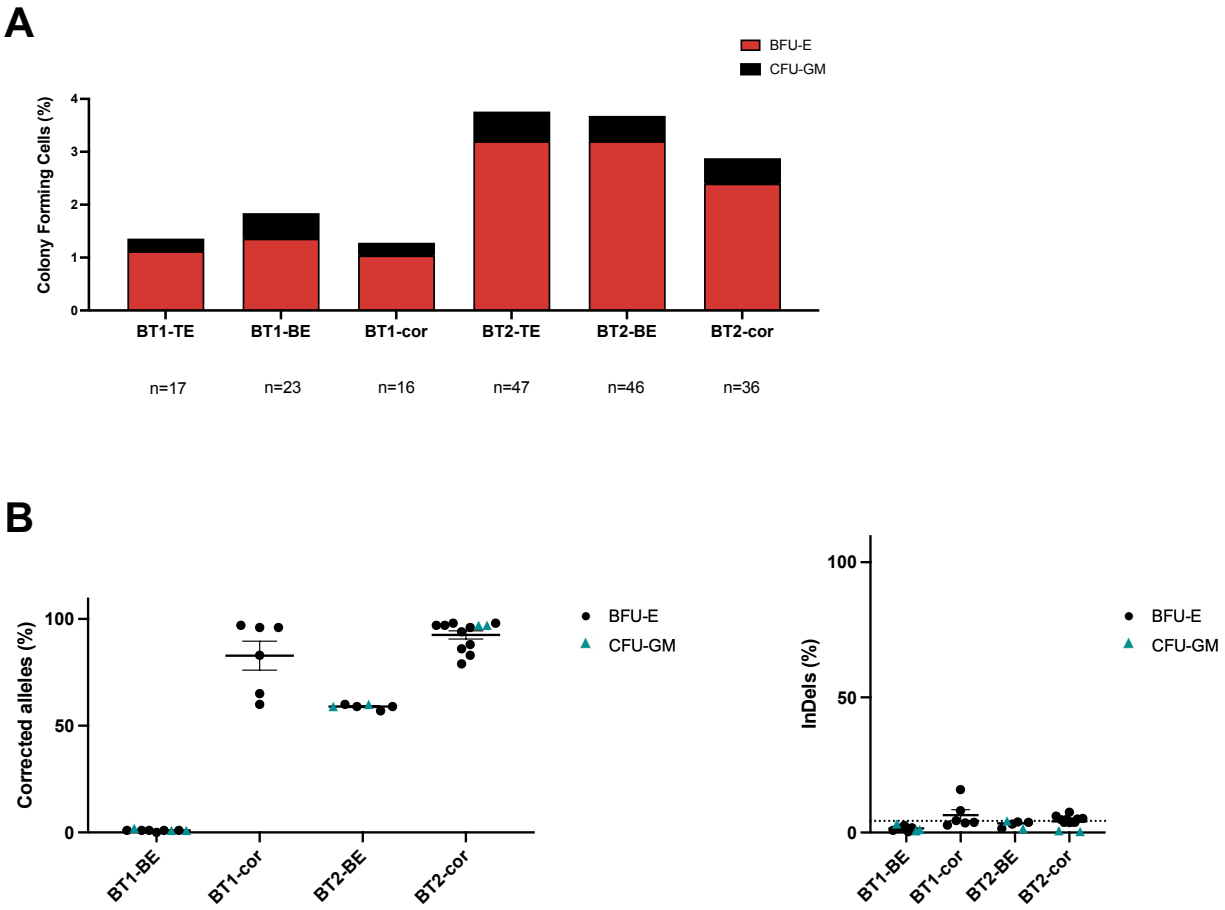


Supplemental Figure 1

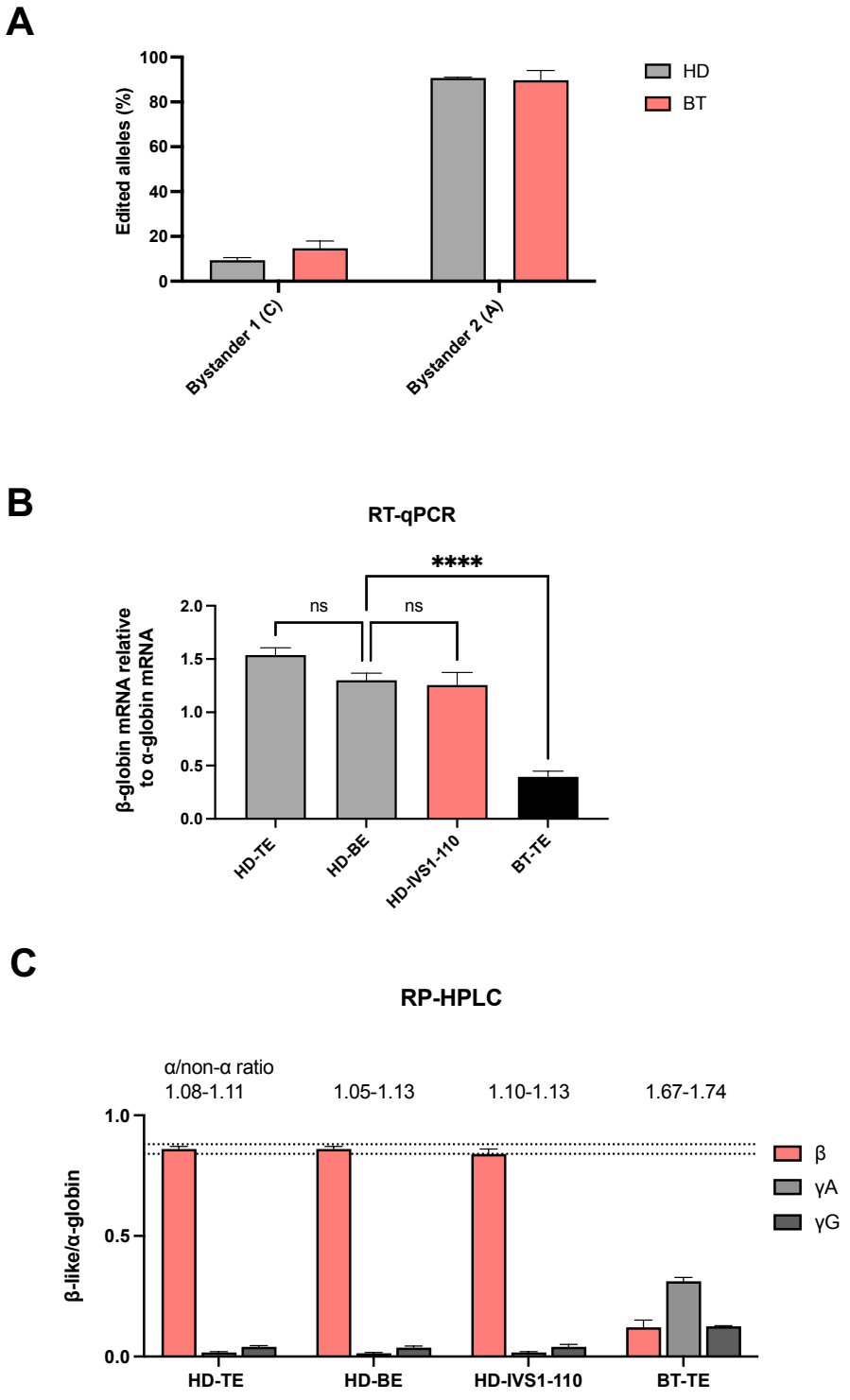




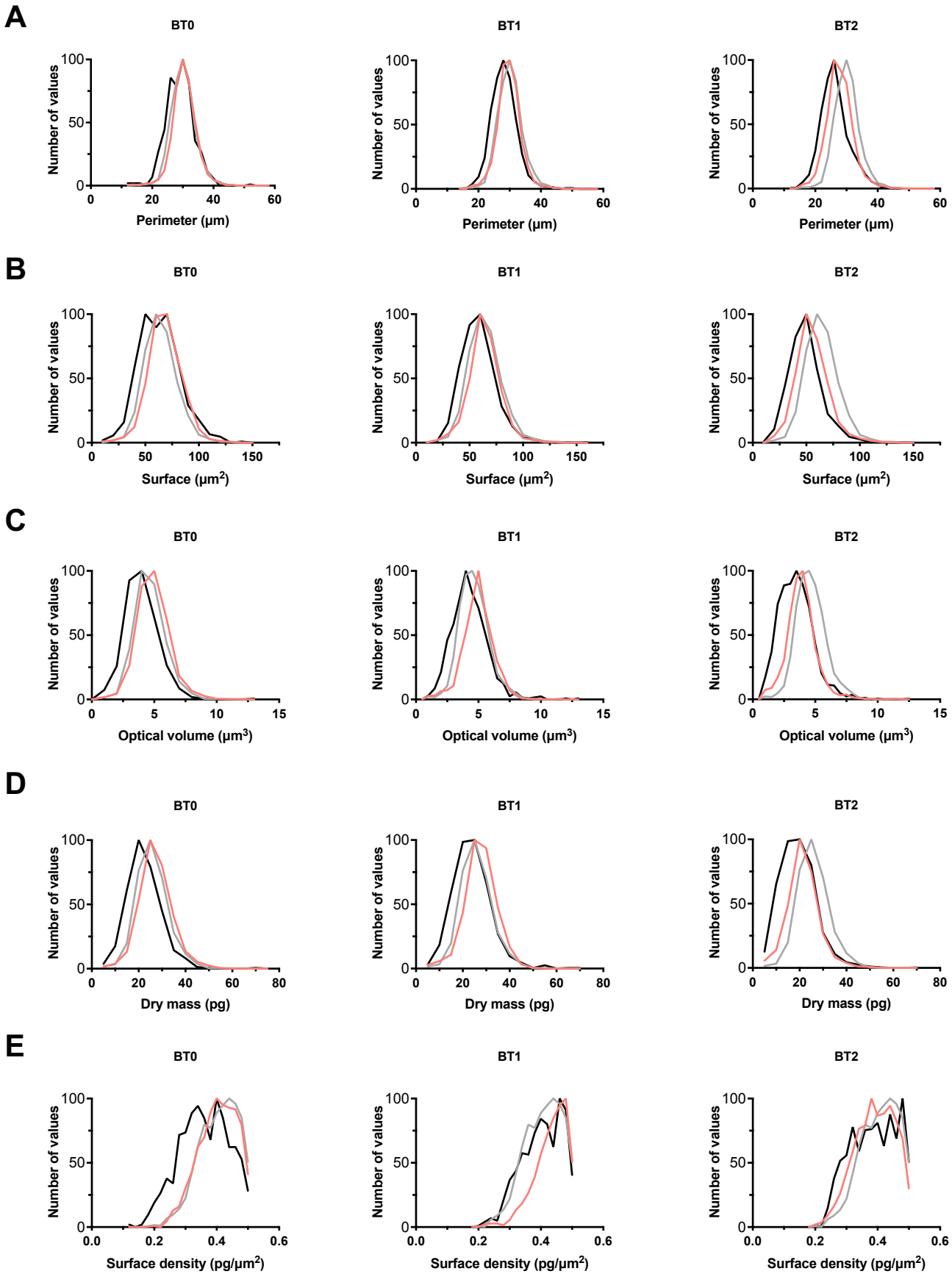
Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4



2. Supplemental figure legends

Supplemental Figure 1. On and off-target editing in β-thalassemic cells.

A. Percent composition of Sanger sequencing traces measured to be significantly different from noise (in red), as assessed by EditR following Sanger sequencing in $CD34^+$ cells transfected with gRNA1/SpRY-ABE8e (cor) or with TE buffer (TE). Target base position is highlighted with a red box and the bystander edits with black boxes (3 donors). **B.** Frequency of base editing (left panel) and InDels (right panel) at the *HBB* gene and at the top-10 predicted off-targets (OTs) in control (BT-TE) and edited (BT-cor) β-thalassemic samples, as measured by targeted NGS sequencing (3 β-thalassemia patients). **** p≤0.0001 (Two-way ANOVA). On-target editing for the compound heterozygous BT2 was calculated considering all the reads harboring a G at the target site as "edited". **C**. Frequency and sequence of modified and unmodified alleles in corrected β-thalassemic samples, as measured by targeted NGS sequencing. Target base position is highlighted with a red box. Bystander edits are present at positions 6 and 8 (black boxes), and at positions -1 and -2 although at low frequency. **D**. Product purity of SpRY-ABE8e, as indicated by the type of substitution measured by targeted NGS sequencing.

Supplemental Figure 2. CFC assay from β-thalassemic HSPCs.

A. CFC frequency in control (TE or BE) and corrected β-thalassemic HSPCs (cor). (2 β-thalassemia patients, BT0 data were not plotted as we obtained only few colonies even in control samples, i.e. ≤10). **B.** Frequency of corrected alleles (left panel) and InDels (right panel), as evaluated by Sanger sequencing in single colonies derived from corrected and control β-thalassemic HSPCs (BT1 and BT2 β-thalassemia patients). In the right panel, the dotted line represents the maximum background noise of InDels calculated by TIDE.

Supplemental Figure 3. Impact of bystander edits on *HBB* expression.

A. Frequency of alleles carrying bystander 1 (C) at position 6 and bystander 2 (A) at position 8 in edited healthy donor (HD) and β-thalassemia (BT) samples (3 β-thalassemia patients and 2 HD). Data are expressed as mean±SEM. **B.** RT-qPCR detecting β-globin mRNAs in erythroid cells derived from edited HD cells carrying the

bystander edits (HD-IVS1-110). As controls, we used erythroid cells derived from HD HSPCs transfected only with TE (HD-TE) or SpRY-ABE8e mRNA (HD-BE). Data from erythroid cells derived from HD (n=2 donors) and control β-thalassemic HSPCs were included (BT-TE; 3 β-thalassemia patients). β-globin expression was normalized to α-globin. Data are expressed as mean±SEM. **** p≤0.0001 (One-way Anova, HD-BE *vs* other conditions). **C.** Expression of β-, ^{G}γ- and ^{A}γ- globin chains measured by RP-HPLC in HD RBCs. β-like-globin expression was normalized to α-globin. The α-/non-α globin ratio is reported on top of the graph. RBCs were obtained from edited HD cells carrying the bystander edits (HD-IVS1-110). As controls, we used erythroid cells derived from HD HSPCs transfected only with TE (HD-TE) or SpRY-ABE8e mRNA (HD-BE). Data from erythroid cells derived from HD (n=2 donors) and β-thalassemic HSPCs were included (BT-TE; 3 β-thalassemia patients). Data are expressed as mean±SEM. Dotted lines indicate maximum and minimum values observed for β-globin in HD-TE conditions.

Supplemental Figure 4. Analysis of RBC parameters by quantitative phase microscopy.

A-E. Quantitative phase microscopy analysis of individual RBCs obtained from corrected β-thalassemic HSPCs (BT-cor, red curve) at day 19 of differentiation. As controls, we used RBCs derived from β-thalassemic HSPCs transfected only with SpRY-ABE8e (BT-BE, black curve) or HD HSPCs transfected only with SpRY-ABE8e (HD-BE, gray curve) (3 β-thalassemia patients and 2 HDs). We reported the relative number of RBCs with different perimeter (A), surface (B), optical volume (C), dry mass (D) and surface density (E) values. Data were reported as overlaid histograms.

3. Supplemental discussion

Our results demonstrate that ABEs can efficiently correct the IVS1-110 mutation in patients' HSPCs without causing DSBs and restore a normal to nearly normal Hb expression profile in HSPC-derived RBCs, thus correcting the β-thalassemic cell phenotype *in vitro*. Targeting only the $\beta^{IVS1\text{-}110(G>A)}$ allele in the compound heterozygote largely corrects ineffective erythropoiesis, but is less effective in ameliorating the α-/non-α-globin ratio than correcting both mutations in

homozygous patients' cells. However, the observed α-/non-α-globin ratios might be close to those observed in erythroid cells from asymptomatic heterozygous carriers of β-thalassemic mutations. If necessary, to fully correct the β-thalassemic phenotype, we envision that the strategy proposed in this study could be combined with an approach either correcting the second mutation or reducing α-globin expression[126].

Other strategies were proposed to specifically target the highly prevalent IVS1-110 mutation[169–171]. Short hairpin RNA-mediated targeting of the aberrant mRNA increases β-globin chain production probably by alleviating translational inhibition. However, our strategy not only reduces the abnormally spliced transcripts but also increases the levels of correctly spliced mRNAs. Compared to nuclease-based strategies disrupting the abnormal splice acceptor site, our approach is more predictable and accurate as all the events lead to correction of the splicing defect.

Other gene therapy strategies for β-thalassemia aim at inducing fetal hemoglobin (HbF). These strategies have the advantage of being potentially applied to all β-thalassemia patients independently from the specific mutation, although it is still debated if high HbF levels are harmful developmentally or physiologically[177]. Furthermore, for prevalent mutations such as IVS1-110, development of a mutation-specific correction strategy could allow the production of higher levels of therapeutic hemoglobin, as *HBB* promoter activity and *HBB* gene expression is favored in adult cells[178]. Finally, our mutation-specific approach counteracts the partially dominant causative effect of aberrant *HBB* mRNAs that is likely still present after HbF induction or expression of a therapeutic globin in LV-based approaches.

4. Supplemental materials and methods

Base editor-expressing plasmids

Plasmids used in this study include NG-ABE8e (Addgene, 138491), ABE8e (Addgene, 138489) and pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP (RTW5025) (Addgene, 140003). The SpRY-ABE8e plasmid was created by replacing the Cas9 sequence of the ABE8e plasmid by the SpRY-P2A-EGFP from the pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP (RTW5025) plasmid.

gRNA design

We manually designed gRNAs targeting the IVS1-110 (G>A) mutation (**Supplemental Table 1**). To evaluate the impact of bystander edits, we adapted gRNA1 sequence to target the WT *HBB* gene (gRNA1_HD). We used chemically modified synthetic gRNAs harboring 2′-O-methyl analogs and 3′-phosphorothioate nonhydrolyzable linkages at the first three 5′ and 3′ nucleotides (Synthego).

mRNA *in vitro* transcription

NG-ABE8e, ABE8e or SpRY-ABE8e plasmids were digested overnight with SapI (Thermo). The linearized plasmids were purified (QIAGEN, 28106) and eluted in DNase/RNase-free water. The *in vitro* transcription protocol (Invitrogen, AM1334) was modified as follows. The GTP nucleotide solution was used at a final concentration of 3 mM and the anti-reverse cap analog N7-Methyl-3'-O-Methyl-Guanosine-5'-Triphosphate-5'-Guanosine (Trilink, N-7003) was added at a final concentration of 12 mM. The incubation time was reduced to 30 minutes and followed by polyadenylation (Invitrogen, AM1350). mRNA was precipitated using LiCl.

Cell purification and culture

We obtained human peripheral blood non-mobilized $CD34^+$ HSPCs from β-thalassemia patients. HD $CD34^+$ HSPCs samples were obtained from bone marrow of healthy donors. Written informed consent was obtained from all subjects. All experiments were performed in accordance with the Declaration of Helsinki. The study was approved by the regional investigational review board (reference: DC 2014-2272, CPP Ile-de-France II “Hôpital Necker-Enfants malades”). HSPCs were purified using the CD34 MicroBead Kit (Miltenyi Biotec). Forty-eight hours before transfection, $CD34^+$ cells were thawed and cultured at a concentration of $5x10^5$ cells/mL in “HSPC medium” containing StemSpan (STEMCELL Technologies) supplemented with penicillin/streptomycin (Gibco), StemRegenin1 (STEMCELL Technologies) and the following recombinant human cytokines (PeproTech): human stem cell factor, Fms-like tyrosine kinase receptor 3 ligand, thrombopoietin and interleukin-3. Four days before transfection, the $CD34^-$ fraction was thawed and cultured at a concentration of $5x10^6$ cells/mL in “T cell medium” containing RPMI 1640 + GlutaMAX (Gibco) supplemented with FBS (Thermo), penicillin/streptomycin

(Gibco) and recombinant human interleukin-2 (Peprotech). Five hours after thawing, cells were transferred to “T cell activation medium” containing RPMI 1640 + GlutaMAX (Gibco) supplemented with CD28 Monoclonal Antibody (eBioscience, Clone CD28.2) in plates coated with CD3 Monoclonal Antibody (eBioscience, Clone OKT3).

RNA transfection

$1x10^6$ T cells or $1x10^4$ to $5x10^5$ HSPCs per condition were transfected with 3.0 µg of the ABE-encoding mRNA and 3.2 µg of the synthetic gRNA using the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the EO-115 or CA-137 program (Nucleofector 4D), respectively. For T cells, when ABE was not fused to GFP (ABE8e, NG-ABE8e), 0.25 µg of GFP-encoding mRNA (Trilink, L-7201) was added to the transfection mix. Cells transfected only with TE buffer or only with ABE-encoding mRNA served as negative controls.

Colony-forming cell (CFC) assay

$CD34^+$ HSPCs were plated at a concentration of 500 cells/mL (according to pre-nucleofection counting) in a methylcellulose-based medium (Stem Cell Technologies, GFH4435) under conditions supporting erythroid and granulo-monocytic differentiation. BFU-E and CFU-GM colonies were counted after 14 days. Single colonies were collected to evaluate base editing efficiency and InDels.

HSPC differentiation

Transfected $CD34^+$ HSPCs were differentiated into mature RBCs using a three-phase erythroid differentiation protocol, as previously described[156,179].

Evaluation of editing efficiency

Genomic DNA was extracted (Invitrogen, K182002 or Zymo Research, ZD7001) from HSPC-derived erythroid cells at day 9 of differentiation. To evaluate base editing efficiency at gRNA target sites, we performed a nested PCR using previously published primers[170], followed by Sanger sequencing and EditR analysis[180]. TIDE analysis (Tracking of InDels by Decomposition) was performed to evaluate the percentage of InDels in edited samples[181].

On- and off-target regions were also PCR-amplified and subjected to NGS. Off-targets were *in silico* predicted using COSMID[182]. We selected the top-10 predicted off-targets and assessed editing at day 9 or 13 of differentiation. On-target and off-target sites were PCR-amplified using the Phusion High-Fidelity polymerase (NEB, M0530) and primers containing specific DNA stretches (MR3 for forward primers and MR4 for reverse primers; **Supplemental Table 3**) located 5' to the sequence recognizing the off-target. For the on-target site, a nested PCR was applied as described above. Amplicons were purified using Ampure XP beads (Beckman Coulter, A63881). Illumina-compatible barcoded DNA amplicon libraries were prepared by a second PCR step using the Phusion High-Fidelity polymerase (NEB, M0530) and primers containing Unique Dual Index (UDI) barcodes and annealing to MR3 and MR4 sequences. Libraries were pooled, purified using the High Pure PCR Product Purification Kit (Sigma-Aldrich, 11732676001), and sequenced using Illumina NovaSeq 6000 system (paired-end sequencing; 2×100-bp) to obtain a minimum of 100,000 reads per amplicon. Targeted NGS data were analyzed using CRISPResso2[183].

RT-PCR and RT-qPCR

RNA was extracted from cells at day 13 of differentiation (Qiagen, 74004 or Zymo Research, ZD7001) and retro-transcribed (Thermo, 18080051). RT-PCR analysis of β-globin mRNAs was performed using previously described primers spanning the exon 1-2 junction[170]. RT-qPCR was performed using previously described primers for the detection of correctly spliced β-globin mRNA[170] and the following primers amplifying α-globin transcripts: α-globin-F 5'-CGGTCAACTTCAAGCTCCTAA-3' and α-globin-R 5'- ACAGAAGCCAGGAACTTGTC-3'. For detection of the effect of bystander edits on *HBB* expression, the following primers were used for RT-qPCR: β-globin-F 5'-GCAAGGTGAACGTGGATGAAGT-3' and β-globin-R 5'-TAACAGCATCAGGAGTGGACAGA-3'.

Flow cytometry analysis

Flow cytometry analysis of erythroid surface markers on HSPC-derived erythroid cells was performed using anti-CD36 (BD Horizon, 561535), anti-CD71 (BD Pharmingen, 555536) and anti-GPA (BD Pharmingen, 563666) antibodies. Flow cytometry analysis of enucleated or viable cells was performed using DRAQ5

(Invitrogen, 65-0880-96) and 7AAD (BD Pharmingen, 559925) dyes, respectively. Apoptosis was evaluated using PE Annexin V Apoptosis Detection Kit I (BD Biosciences). Flow cytometry analysis of reactive oxygen species (ROS) was performed using H_2DCFDA (Invitrogen, D399). Flow cytometry analyses were performed using Gallios (Beckman coulter) flow cytometer. Data were analyzed using the FlowJo (BD Biosciences) software.

RP-HPLC and CE-HPLC

HPLC analysis was performed in HSPC-derived erythroid cells at day 16 of differentiation as previously described[156].

Quantitative phase image microscopy of RBCs

180,000 *in vitro* differentiated RBCs (at day 19 or 20) were resuspended in CellStab (BioRad, 005650) and placed in an 8-well m-Slide (Ibidi, 80826). We used the SID4 HR GE camera (Phasics) with an inverted microscope (Nikon, Eclispe Ti-E) to obtain quantitative phase images of label-free RBCs. The BIO-Data R&D software (version 2.7.1.46) was used to perform a segmentation procedure and isolate each RBC. We analyzed only enucleated RBCs, discarding cells with a surface density (dry mass/surface) higher than 0.500 pg/μm^2.

5. Supplemental Tables

Supplemental Table 1. gRNA target sequences.

gRNA	Sequence (5' to 3')	PAM sequence (5' to 3')	Position (hg19)	Strand	Compatible ABEs
gRNA1	TTAGTCTATTTTCCCACCCT	TAG	chr11:5248033-5248052 (*HBB*)	-	SpRY-ABE8e, NG-ABE8e, ABE8e
gRNA2	ATTAGTCTATTTTCCCACCC	TTA	chr11:5248034-5248053 (*HBB*)	-	SpRY-ABE8e
gRNA3	TATTAGTCTATTTTCCCACC	CTT	chr11:5248035-5248054 (*HBB*)	-	SpRY-ABE8e
gRNA4	CTATTAGTCTATTTTCCCAC	CCT	chr11:5248036-5248055 (*HBB*)	-	SpRY-ABE8e
gRNA5	CCTATTAGTCTATTTTCCCA	CCC	chr11:5248037-5248056 (*HBB*)	-	SpRY-ABE8e
gRNA6	GCCTATTAGTCTATTTTCCC	ACC	chr11:5248038-5248057 (*HBB*)	-	SpRY-ABE8e
gRNA1_HD	TTGGTCTATTTTCCCACCCT	TAG	chr11:5248033-5248052 (*HBB*)	-	SpRY-ABE8e, NG-ABE8e, ABE8e

PAM, Protospacer Adjacent Motif.

Supplemental Table 2. *In silico* predicted off targets.

Off-target	Sequence (5' to 3')	Mismatches	Position (hg19)	Strand	Score	Type
OT1	GTAGTCTATTTTCCCACCCT	1	Chr10:36827935-36827954	+	0.12	Intergenic
OT2	CTAGCCTATTTTCCCACCCT	2	Chr1:104536228-104536247	-	0.31	Intergenic
OT3	TCAGTCCATTTTCCCACCCT	2	ChrX:136953484-136953503	+	0.36	Intergenic
OT4	TTGGTCGATTTTCCCACCCT	2	Chr8:129251336-129251355	+	0.38	Intergenic
OT5	TTAGGTTATTTTCCCACCCT	2	Chr3:65005077-65005096	-	0.40	Intergenic
OT6	TTTGTCTCTTTTCCCACCCT	2	Chr12:40609156-40609175	+	0.42	Intergenic
OT7	TTTGTCTCTTTTCCCACCCT	2	Chr8:89004815-89004834	-	0.42	Intergenic
OT8	TTAGTAAATTTTCCCACCCT	2	Chr12:11337215-11337234	+	0.44	Intronic – *PRB4*
OT9	TTAGCCTAGTTTCCCACCCT	2	ChrX:142880716-142880735	+	0.54	Intergenic
OT10	TTAGCTATTTTCCCACCCT	0 (Del 16)	Chr4:59423629-59423647	-	0.70	Intergenic

OT, Off-target; Del, Deletion.

Supplemental Table 3. Primers used for targeted NGS.

Amplified region	F/R	Sequence (5' to 3')
On target	F	GCAGCGTCAGATGTGTATAAGAGACAGCTGGGCATGTGGAGACAGAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTTGCCATGAGCCTTCACCTT
OT1	F	GCAGCGTCAGATGTGTATAAGAGACAGGCTCAGAAACAGGATGTCATCATTTCCTC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTTCCCTAAATCAGCAACCCTTGG
OT2	F	GCAGCGTCAGATGTGTATAAGAGACAGGGTCATATTACAGGCAAAAATAGAAAATCCTTGTGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGAAAAGTGCAACTTGGTGGCAGAATCTAATTTGAATC
OT3	F	GCAGCGTCAGATGTGTATAAGAGACAGTGTTGGATTCTTGCAGGCACGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGGATACAGAGAGAGGATGGAGAG
OT4	F	GCAGCGTCAGATGTGTATAAGAGACAGGATGCTAGGATCTGACTTCTGTGC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCCAAGAGCACTAGCCCTTCAC
OT5	F	GCAGCGTCAGATGTGTATAAGAGACAGCCCTCGGGCTTTTAACACATGC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCAGCTTCCACATGAGATGATTCCC
OT6	F	GCAGCGTCAGATGTGTATAAGAGACAGCTGGATCCCTTCAATAGCCTCCT
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGTGGCCACTGCAAAGTCTTCC
OT7	F	GCAGCGTCAGATGTGTATAAGAGACAGCATCTTCAGTAAGTCAGTGTAGATGTACTTTCC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTTTTTTCCCTCTTCTGCCTCCTCC
OT8	F	GCAGCGTCAGATGTGTATAAGAGACAGCAGCTTCTGTACAGTGTTTTAATCTATGCCATC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGACAAGCATATGTATACATAGTGCTGCTTTGAC
OT9	F	GCAGCGTCAGATGTGTATAAGAGACAGTGTCTCAGGCTCCCAAGCTTC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCAGACTTGGTGGATTTTACTAAGACCCG
OT10	F	GCAGCGTCAGATGTGTATAAGAGACAGTGGGCCATTCAATCTTTACTTTTG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGGAGTGGTTAACTGCCCCATT

F, forward primer; R, reverse primer.

IV. Base editing approaches for the treatment of two β^0 thalassemia mutations.

A. Introduction

β-globin chains together with α-globin chains compose the adult hemoglobin (HbA) tetramer. β-thalassemia is a monogenic recessive disease caused by a variety of mutations affecting the synthesis of the adult hemoglobin β-chains. It is a highly prevalent hemoglobinopathy with 68,000 affected children annually worldwide[164]. Patients were originally concentrated in Asia, India and the Mediterranean region but due to recent population movements, β-thalassemia is now widely spread in Europe and North America[164]. Point mutations or deletions in the β-globin gene (*HBB*) locus reduce (β^+) or abolish (β^0) the production of β-globin chains. The imbalance between α- and β-globin production, leads to the precipitation of uncoupled α-globins, which causes erythroid cell death, ineffective erythropoiesis and severe anemia[184]. Depending on their genotype, the clinical phenotype can vary from mild to severe anemia, known as β-thalassemia major (typically associated with a β^0/β^0 genotype). In the most severe cases, patients are transfusion-dependent and require an iron-chelation therapy to alleviate iron overload due to chronic transfusions.

The only definitive cure for β-thalassemia patients is the transplantation of allogeneic hematopoietic stem cells (HSCs), but this treatment is limited by the availability of HLA-compatible donors. Gene therapy approaches based on the transplantation of autologous, genetically modified HSCs have been investigated as a treatment option for patients lacking a compatible donor[2]. In particular, genome editing technology has been recently exploited to develop therapeutic approaches for β-thalassemia[112,144,157].

CD39 (CAG>TAG) (hereinafter referred to as “CD39”) is one of the most common β^0-thalassemic mutation in the Mediterranean area and Latin America, representing >40% of β-thalassemic mutations in Tunisia, Argentina and Italy[185]. This is a nonsense mutation within the codon 39, thus it causes premature translation termination and absence of β-globin[186]. On the other hand, IVS2-1 (G>A) (hereinafter referred to as “IVS2-1”) is one of the most common β^0-thalassemic mutations in the Middle East, representing around 30% of β-thalassemic mutations found in Iran and Kuwait[187]. This point mutation occurs at the first nucleotide of the second intron of *HBB* and disrupts the splice donor site. It results in the production of two abnormally

spliced mRNAs[188]. The first mRNA is characterized by a partial retention of the second intron, whereas the second mRNA consists of the first exon spliced directly with the third one. Importantly, none of them lead to the production of β-globin.

Genome editing approaches to correct CD39 and IVS2-1 mutations have already been developed. Approaches stimulating gene correction in the presence of a WT single-stranded DNA donor (containing a wild-type β-globin gene fragment) through peptide nucleic acid (PNA) binding enabled precise correction of the IVS2-1 mutation, but with very low frequencies (<1%)[189]. Other approaches use designer nucleases, such as the CRISPR/Cas9 nuclease system that induces DNA double-strand breaks (DSBs) via a single guide RNA (gRNA) complementary to a specific genomic target. The DSB can be repaired via homologous-directed repair (HDR), by providing a donor DNA template containing the wild type sequence, allowing direct gene correction of the mutation. An HDR-based strategy correcting the CD39 mutation was tested in human erythroid precursors and in hematopoietic progenitor cells reaching very low levels of correction (~5% and ~10% of the alleles)[190]. In *bona fide* HSCs, the target cell population in gene therapy, HDR-mediated gene correction is known to be poorly efficient[66] and even less efficient than in hematopoietic progenitor cells. Furthermore, HSCs are highly sensitive to DSBs[191] - especially in cases of multiple on-targets or concomitant on-target and off-target events. Even when highly specific gRNAs are used, Cas9-gRNA treatment of human hematopoietic/stem progenitor cells (HSPCs) induces a DNA damage response that can lead to apoptosis[192,193]. CRISPR/Cas9 can cause p53-dependent cell toxicity and cell cycle arrest, resulting in the negative selection of cells with a functional p53 pathway[172]. Furthermore, the generation of several on-target DSBs, simultaneous on-target and off-target DSBs, or even a single on-target DSB is associated with a risk of deletion, inversion and translocation[194–197].

Base editing, a new CRISPR/Cas9-derived genome editing tool, allows precise DNA repair in *bona fide* HSCs[126] without the occurrence of DSBs. Adenine base editors (ABE) and cytosine base editors (CBE) contain a Cas9 nickase and a deaminase, and permit the insertion of A•T > G•C and C•G >T•A mutations, respectively[75]. Base editing using CBEs has already been exploited to correct a β-thalassemia-causing mutation in the *HBB* promoter[89,126]. Here, we used ABEs to precisely correct the

CD39 and IVS2-1 mutations in patient HSCs and demonstrated the rescue of the thalassemic phenotype in their erythroid progeny.

B. Results

Adenine base editing is an efficient tool to correct the CD39 (CAG>TAG) and IVS2-1 (G>A) mutations

ABEs allow A•T > G•C conversions and can potentially correct the CD39 (CAG>TAG) and IVS2-1 (G>A) mutations. Therefore, we manually designed gRNAs placing the target base within the canonical editing window of ABEs, from positions 4 to 8 (**Figure 1A**). For IVS2-1 mutation, we designed gRNAs that do not target a SNP commonly found in β-thalassemic patient, thus limiting the number of gRNAs to 3[198,199] (**Figure 1A**). All the gRNAs were tested in combination with SpRY-ABE8e, and, if compatible, with ABE8e combined with nCas9s recognizing more restrictive PAMs, such as NG or NRCH (**Table 1**).

We screened gRNA/BE combinations in T cells obtained from β-thalassemic patients homozygous for the CD39 and IVS2-1 mutations (**Figure 1B**, BT0 and BT3 patients). Cells were transfected with chemically modified gRNAs and *in vitro* transcribed ABE mRNAs. gRNA1/SpRY-ABE8e and gRNA1/NRCH-ABE8e were the two most efficient combinations able to revert the CD39 mutation in more than 90% of *HBB* alleles, as evaluated by Sanger sequencing and EditR analysis (**Figure 1C**). On the other hand, gRNA2/SpRY-ABE8e was the most efficient combination able to revert the IVS2-1 mutation, correcting more than 50% of the alleles (**Figure 1C**). In all the cases, we observed low or negligible InDel generation (**Figure 1C**).

Efficient correction of the CD39 and IVS2-1 mutations in β-thalassemic HSPCs restores normal Hb production in their erythroid progeny

HSPCs from β-thalassemia patients were transfected with selected chemically modified gRNAs and *in vitro* transcribed ABE mRNAs (**Figures 1B, 2A**). For CD39, we used the NRCH-ABE8e enzyme as a more restrictive PAM requirement is expected to produce less off-targets effects. One donor was homozygous for the CD39 mutation (BT0) and other two, BT1 and BT2, were compound heterozygotes harboring the CD39 mutation and another β^0 or β^+ mutation, respectively (**Figure 1B**). For evaluating the correction of the IVS2-1 mutation, HSPCs were harvested

from one homozygous patient (**Figure 1B**, BT4). Deep sequencing of edited samples demonstrated that the correction of the targeted mutation was highly efficient and reproducible among the replicates (CD39: 98.1%±0.5, and IVS2-1: 90.4%±1.5) (**Figure 2B**). Moreover, it confirmed the DSB-free nature of ABEs, as we detected low to no InDels in base-edited samples (**Figure 2B**).

Of note, several bystanders were observed close to the CD39 mutation (**Figure 2B**, **Supplementary Figure 1**, b0 to b4). While correction of the mutation restores the CAG codon specifying glutamate, editing of an adjacent cytosine in parallel can either lead to a non-synonymous mutation (b3 = CAG>CAC: Glu>His) or to a synonymous mutation (CAG>CAA) (**Figure 2B**, **Supplementary Figure 1**, b3). Importantly, this Glu>His amino acid change (occurring in ~4% of total alleles) has likely no consequences as it was described in a known Hb variant (Hb San Bruno), which is not associated with any clinical or hematological abnormalities[200]. The four other bystander edits (all occurring in <1% of total alleles) generated non-synonymous mutations (b4 = TGG>CGG: Trp>Arg, also known as CD37; b2 = AGG>AAG: Arg>Lys, also known as CD40; b1 = TTC>CTC: Phe>Leu or b0 = TTC>TCC: Phe>Ser, also known as CD41) leading to previously described Hb variants not associated with any hematological feature or associated to a mild cyanosis (for the CD41 (Phe>Ser) mutation and for the CD37 (Trp>Arg) one) (**Supplementary Figure 1**, b0, b1, b2 and b4 bystanders)[201–205].

In the case of IVS2-1, two bystander mutations were observed (**Figure 2B, Supplementary Figure 1**, b0 and b1). The first one maps to the last codon of *HBB* exon 1 (**Figure 2B, Supplementary Figure 1**, b0). This event leads to a missense mutation, replacing an arginine by a glycine. When combined with the correction of the splicing mutation (in ~73% of corrected alleles), it will lead to the production of a known Hb variant (Hb-Nîmes), which is not associated with any clinical abnormalities[206]. The second bystander edit (~6% of total alleles) corresponds to an intronic known SNP (rs1468286413) with no reported clinical manifestation[207] (**Figure 2B, Supplementary Figure 1**, b1).

Following transfection, β-thalassemic HSPCs were differentiated towards the erythroid lineage (**Figure 2A**). β-globin mRNA levels in CD39 edited samples were

similar to those observed in HD cells for the homozygous BT0 donor, while representing up to 50% of the HD β-globin transcripts for the compound β^0/β^0 heterozygote and up to 80% for the compound heterozygous β^0/β^+ donor (**Figure 2C**). In IVS2-1 samples, RT-PCR showed restoration of the correct β-globin splicing and a substantial decrease of aberrant transcripts characterized by lower mobility due to partial intron retention or by higher mobility due to exon 2 skipping (**Figure 2D**). The levels of correctly spliced β-globin mRNA were similar to those observed in HD cells in corrected IVS2-1 cells (**Figure 2C**). Of note, primers recognizing the correctly spliced mRNA were designed to hybridize to the exon2-3 junction and thus amplification of the cDNA might be impaired by the creation of the Hb-Nîmes mutation and underestimate the real mRNA amount in corrected samples.

In untreated β-thalassemic RBCs, HPLC showed elevated α-/non-α-globin ratios and the low β-globin expression was poorly compensated by fetal γ (γA+γG)-globins (**Figure 2E**). After treatment, RBCs exhibited higher levels of β-globin chain and HbA (**Figure 2E, F**). For IVS2-1 conditions, we detected Hb-Nîmes representing ~64% of total hemoglobins (**Figure 2F**). Importantly, the α-/non-α-globin ratio was substantially ameliorated in the erythroid cells obtained from corrected β-thalassemic HSPCs (**Figure 2E**).

In conclusion, we were able to efficiently correct the CD39 and IVS2-1 mutations in β-thalassemic HSPCs without causing DSBs, and to restore a therapeutic Hb expression profile in HSPC-derived RBCs.

Efficient reversion of the CD39 (CAG>TAG) and IVS2-1 (G>A) mutations in β-thalassemic HSPCs corrects ineffective erythropoiesis.

In β-thalassemia, α- and β-globin chain imbalance causes premature death via apoptosis of erythroid precursors, thus leading to ineffective erythropoiesis, a hallmark of the disease[176]. The typical delayed erythroid differentiation of β-thalassemic cells was corrected by our treatment, as evaluated by the flow cytometry analysis of different erythroid markers. Indeed, the early erythroid markers CD36 and CD71, were properly downregulated at the end of the differentiation in samples derived from edited HSPCs, similarly to healthy donor samples (**Figure 3A**).

Moreover, in all the samples, we observed an increased enucleation rate (frequency of DRAQ5$^-$ cells) along the differentiation compared to controls, reaching enucleation levels similar to the frequencies observed in healthy donor erythroid populations (**Figure 3B**). At the end of the differentiation, the size of enucleated cells (typically reduced in β-thalassemic cells in culture) was increased in edited samples (**Figure 3C**). The production of ROS observed in β-thalassemic control erythroid cells was significantly decreased in treated samples (**Figure 3D**). Finally, we evaluated the potential of our treatment to rescue the apoptosis in β-thalassemic cells. Flow cytometry showed rescue of edited β-thalassemic erythroid precursors from apoptosis (**Figure 3E**).

C. Discussion

Our results demonstrate that ABEs can efficiently correct the CD39 and IVS2-1 β^0 mutations in patients' HSPCs without causing damaging DSBs. While nonsense-mediated decay is known to rapidly degrade β^0 CD39 carrying β-globin mRNAs, our results show the restoration of normal levels of functional mRNAs[208]. Additionally, reversion of the β^0 IVS2-1 mutation leads to the correction of the splicing defect and production of normal levels of correctly spliced functional mRNAs. These strategies restored Hb expression in HSPC-derived RBCs, thus correcting the β-thalassemic cell phenotype *in vitro*.

Correction of only one β^0 allele in compound heterozygote cells leads to sufficient β-globin expression to largely correct ineffective erythropoiesis, probably by reaching levels of expression close to those observed in erythroid cells from asymptomatic heterozygous carriers of β-thalassemic mutations. If necessary, to fully correct the β-thalassemic phenotype, we envision that the strategy proposed in this study could be combined with an approach either correcting the second mutation or reducing α-globin expression[126].

Other strategies were proposed to specifically correct these mutations. However, they rely on CRISPR/Cas9 or PNA mediated HDR, and were thus limited by the low efficiency and the potential CRISPR/Cas9-induced toxicity[189,190]. On the other hand,

gene therapy strategies aiming at inducing HbF demonstrated their efficacy in correcting the β-thalassemic phenotype[126,157,209] (**Annex 1**). These strategies have the advantage of being potentially applied to all β-thalassemia patients independently from the specific mutation, although it is still debated if high HbF levels in pregnant women impact adversely fetal development or normal physiology[177]. Furthermore, for prevalent and severe mutations such as CD39 and IVS2-1, development of a mutation-specific correction strategy could allow the production of higher levels of therapeutic hemoglobin, as *HBB* promoter activity and *HBB* gene expression is favored in adult cells[178]. In fact, comparison of both strategies by Zeng et al. (HbF induction *vs HBB* β^+ mutation correction) revealed that they both led to similar levels of therapeutic Hb (HbF or HbA) and phenotypic correction but the *HBB* gene correction efficiency was much lower than the frequency of the HbF-inducing mutation[126].

In conclusion, we demonstrated that reverting the CD39 and IVS2-1 β^0 mutations using base editing corrects *in vitro* the β-thalassemic cell phenotype. Validation of these results in HSPCs *in vivo* and genotoxicity studies will provide sufficient proof of efficacy and safety to enable the clinical development of base-edited HSPCs for gene therapy of β-thalassemia.

D. Figures

Figure 1

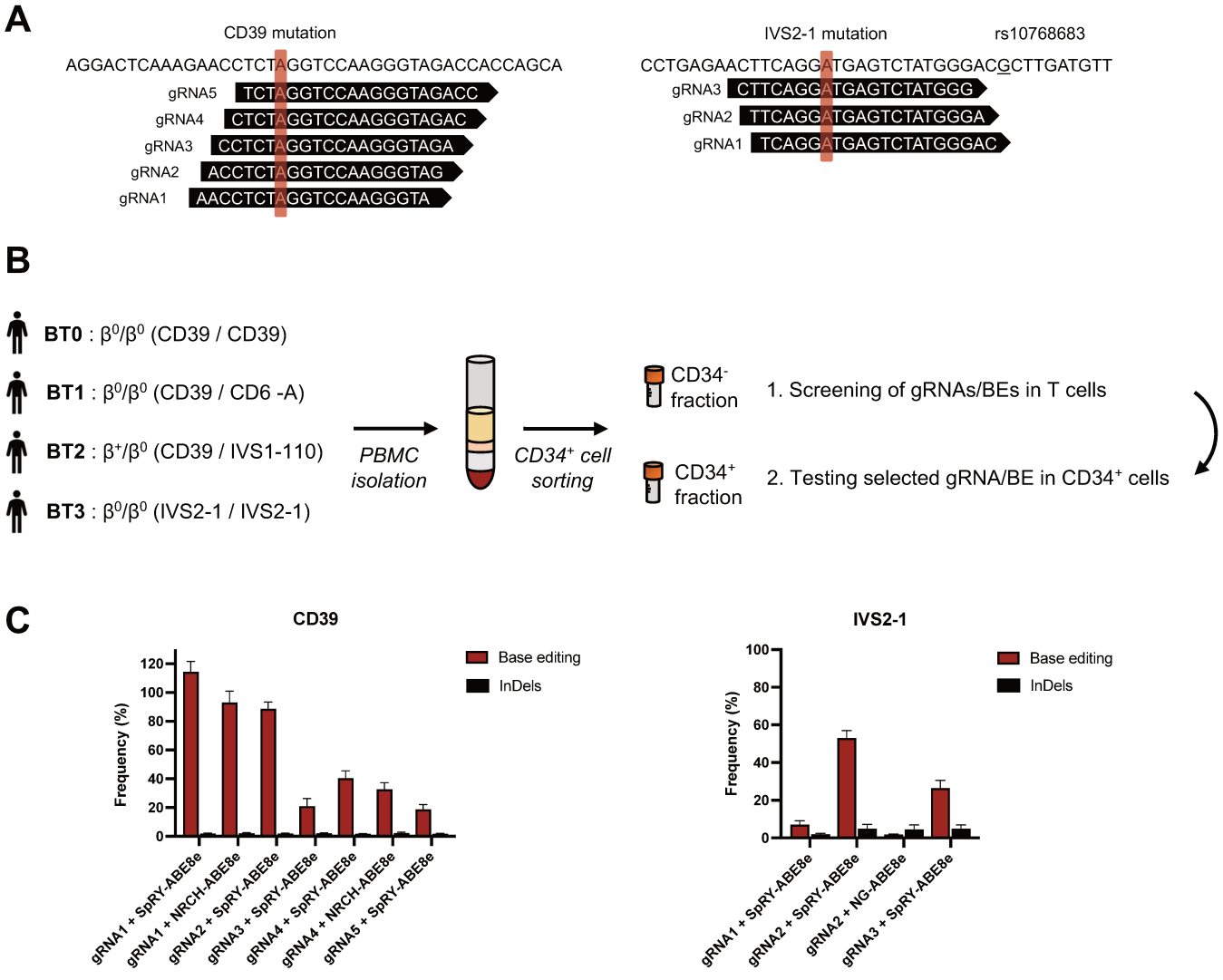

A
CD39 mutation
AGGACTCAAAGAACCTCTAGGTCCAAGGGTAGACCACCAGCA
gRNA5 TCTAGGTCCAAGGGTAGACC
gRNA4 CTCTAGGTCCAAGGGTAGAC
gRNA3 CCTCTAGGTCCAAGGGTAGA
gRNA2 ACCTCTAGGTCCAAGGGTAG
gRNA1 AACCTCTAGGTCCAAGGGTA
IVS2-1 mutation
rs10768683
CCTGAGAACTTCAGGATGAGTCTATGGGACGCTTGATGTT
gRNA3 CTTCAGGATGAGTCTATGGG
gRNA2 TTCAGGATGAGTCTATGGGA
gRNA1 TCAGGATGAGTCTATGGGAC
B
BT0 : β0/β0 (CD39 / CD39)
BT1 : β0/β0 (CD39 / CD6 -A)
BT2 : β+/β0 (CD39 / IVS1-110)
BT3 : β0/β0 (IVS2-1 / IVS2-1)
PBMC isolation
CD34+ cell sorting
CD34- fraction
CD34+ fraction
1. Screening of gRNAs/BEs in T cells
2. Testing selected gRNA/BE in CD34+ cells
C
CD39
Base editing
InDels
Frequency (%)
0
20
40
60
80
100
120
gRNA1 + SpRY-ABE8e
gRNA1 + NRCH-ABE8e
gRNA2 + SpRY-ABE8e
gRNA3 + SpRY-ABE8e
gRNA4 + SpRY-ABE8e
gRNA4 + NRCH-ABE8e
gRNA5 + SpRY-ABE8e
IVS2-1
Base editing
InDels
Frequency (%)
0
20
40
60
80
100
gRNA1 + SpRY-ABE8e
gRNA2 + SpRY-ABE8e
gRNA2 + NG-ABE8e
gRNA3 + SpRY-ABE8e

Figure 2

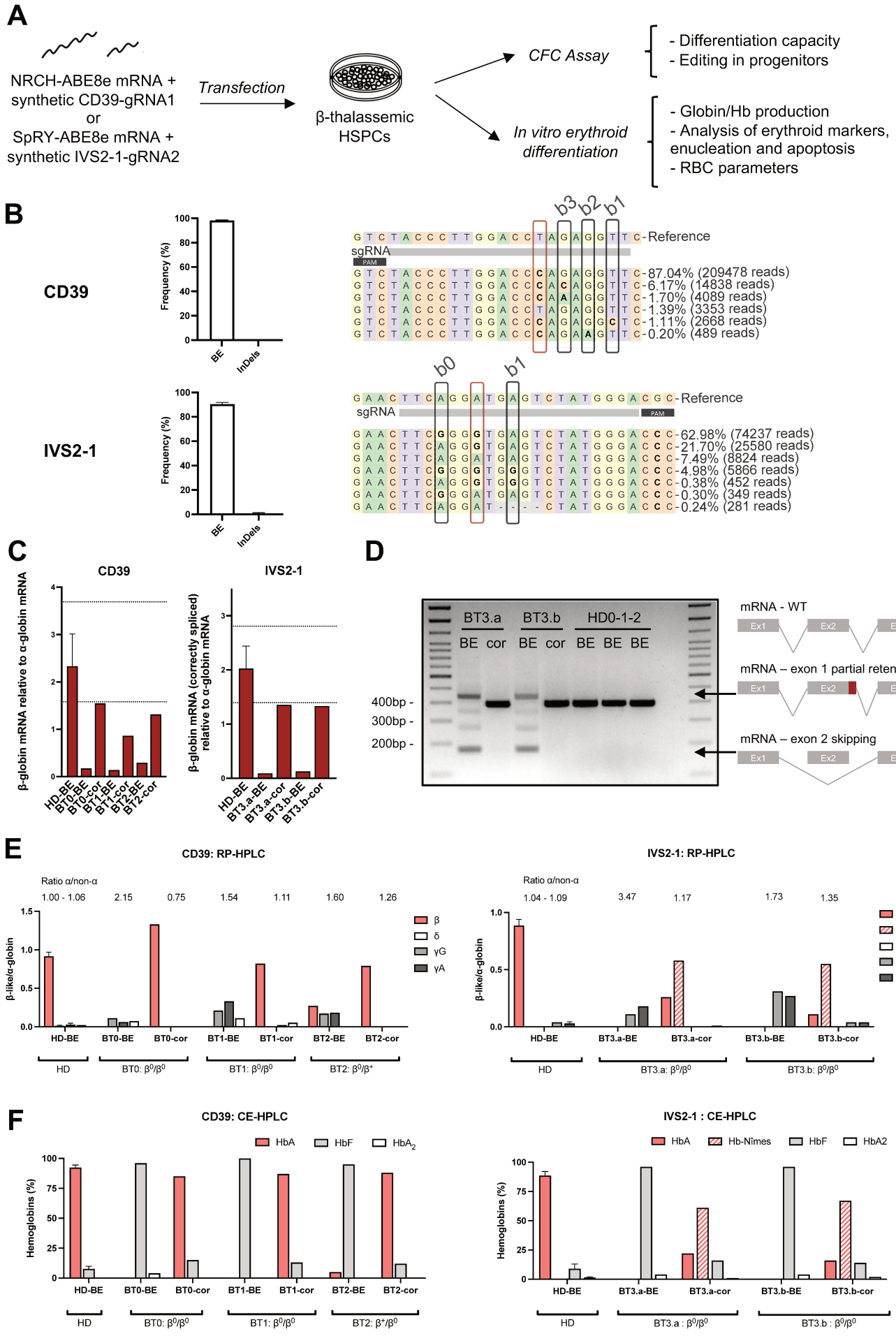
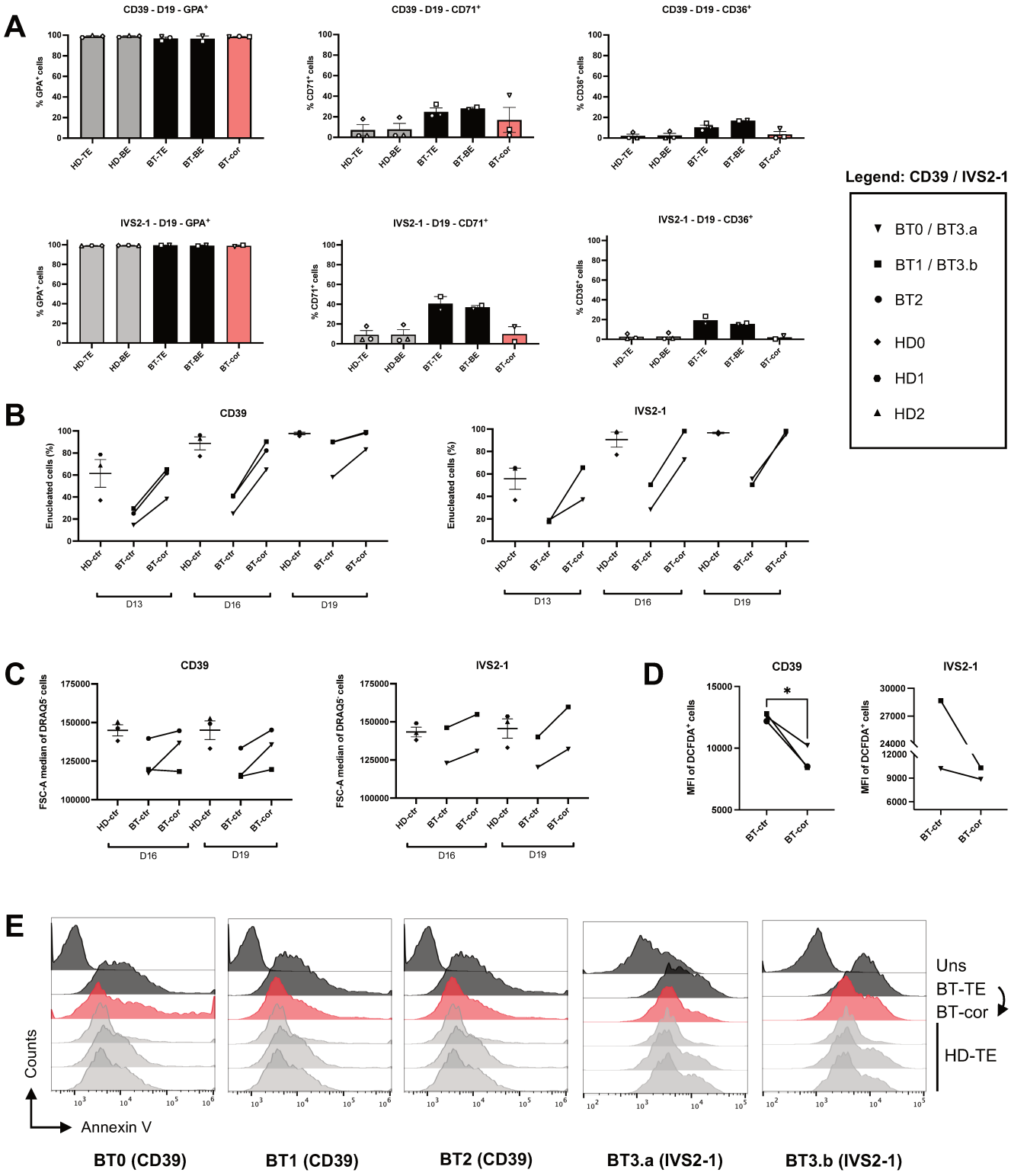

A
NRCH-ABE8e mRNA +
synthetic CD39-gRNA1
or
SpRY-ABE8e mRNA +
synthetic IVS2-1-gRNA2
Transfection
β-thalassemic
HSPCs
CFC Assay
- Differentiation capacity
- Editing in progenitors
In vitro erythroid
differentiation
- Globin/Hb production
- Analysis of erythroid markers,
enucleation and apoptosis
- RBC parameters
B
CD39
IVS2-1
Frequency (%)
BE
InDels
b3
b2
b1
b0
Reference
sgRNA
PAM
87.04% (209478 reads)
6.17% (14838 reads)
1.70% (4089 reads)
1.39% (3353 reads)
1.11% (2668 reads)
0.20% (489 reads)
62.98% (74237 reads)
21.70% (25580 reads)
7.49% (8824 reads)
4.98% (5866 reads)
0.38% (452 reads)
0.30% (349 reads)
0.24% (281 reads)
C
CD39
IVS2-1
β-globin mRNA relative to α-globin mRNA
β-globin mRNA (correctly spliced) relative to α-globin mRNA
HD-BE
BT0-BE
BT0-cor
BT1-BE
BT1-cor
BT2-BE
BT2-cor
BT3.a-BE
BT3.a-cor
BT3.b-BE
BT3.b-cor
D
BT3.a
BT3.b
HD0-1-2
BE cor BE cor BE BE BE
400bp -
300bp -
200bp -
mRNA - WT
mRNA – exon 1 partial retention
mRNA – exon 2 skipping
Ex1
Ex2
Ex3
E
CD39: RP-HPLC
IVS2-1: RP-HPLC
Ratio α/non-α
1.00 - 1.06
2.15
0.75
1.54
1.11
1.60
1.26
1.04 - 1.09
3.47
1.17
1.73
1.35
β-like/α-globin
β
β-Nîmes
δ
γG
γA
HD
BT0: β0/β0
BT1: β0/β0
BT2: β+/β0
BT3.a: β0/β0
BT3.b: β0/β0
F
CD39: CE-HPLC
IVS2-1 : CE-HPLC
HbA
HbF
HbA2
Hb-Nîmes
Hemoglobins (%)
BT2: β+/β0
BT3.a : β0/β0
BT3.b : β0/β0

Figure 3


A
CD39 - D19 - GPA+
CD39 - D19 - CD71+
CD39 - D19 - CD36+
IVS2-1 - D19 - GPA+
IVS2-1 - D19 - CD71+
IVS2-1 - D19 - CD36+
% GPA+ cells
% CD71+ cells
% CD36+ cells
HD-TE
HD-BE
BT-TE
BT-BE
BT-cor
Legend: CD39 / IVS2-1
BT0 / BT3.a
BT1 / BT3.b
BT2
HD0
HD1
HD2
B
CD39
IVS2-1
Enucleated cells (%)
HD-ctr
BT-ctr
BT-cor
D13
D16
D19
C
CD39
IVS2-1
FSC-A median of DRAQ5+ cells
D16
D19
D
CD39
IVS2-1
MFI of DCFDA+ cells
BT-ctr
BT-cor
*
E
Counts
Annexin V
Uns
BT-TE
BT-cor
HD-TE
BT0 (CD39)
BT1 (CD39)
BT2 (CD39)
BT3.a (IVS2-1)
BT3.b (IVS2-1)

E. Figure legends

Figure 1. Design and screening of gRNAs targeting the CD39 (CAG>TAG) and IVS2-1 (G>A) mutations in β-thalassemic T cells.

A. CD39 gRNAs 1-5 and IVS2-1 gRNAs 1-3 were manually designed to place the mutations within the editing window, while avoiding to place known SNPs within the protospacer sequence. The mutation position is highlighted with a red box. SNP rs10768683 position proximal to the IVS2-1 mutation is underlined. **B.** Overview of the cell collection process for testing the ability of gRNA/BE combinations to revert the CD39 and IVS2-1 mutations. PBMCs were isolated from one homozygous and two compound heterozygous β-thalassemia patients harboring the CD39 mutation and 1 homozygous patient harboring the IVS2-1 mutation. After $CD34^+$ cell sorting, T cells were recovered from the negative fraction for testing gRNA/BE combinations, before moving to $CD34^+$ cells with a selected strategy. **C.** Frequency of corrected alleles (normalized to the frequency of GFP^+ cells) as evaluated by EditR and InDel frequency as assessed by TIDE in T cells transfected with different combinations of synthetic gRNAs and ABE mRNAs. Data are expressed as mean±SEM (for each mutation: n=3 biologically independent experiments, 1 homozygous donor).

Figure 2. Efficient correction of the CD39 (CAG>TAG) and IVS2-1 (G>A) mutations in β-thalassemic HSPCs restores normal Hb production.

A. Experimental protocol used for base editing experiments in β-thalassemic HSPCs. NRCH-ABE8e mRNA and synthetic gRNA 1 or SpRY-ABE8e mRNA and synthetic gRNA 2 were co-transfected in β-thalassemic HSPCs harboring the CD39 or the IVS2-1 mutation, respectively. Cells were differentiated into mature RBCs using a three-phase erythroid differentiation protocol or plated in a semi-solid medium favoring hematopoietic progenitors' differentiation. **B.** Frequency of corrected alleles and InDel frequency as calculated after deep sequencing of *HBB* in edited β-thalassemic HSPCs. Data are expressed as mean±SEM (CD39: n=2 biologically independent experiments, 3 donors; IVS2-1: n=2 biologically independent experiments, 1 donor). Frequency of corrected alleles in the cells from compound

heterozygous patients (BT1 and BT2) were corrected to take into account only alleles harbouring the CD39 mutation. Representative NGS reads in corrected samples from homozygous patients carrying CD39 and IVS2-1 mutations are reported in the right panels. Mutation positions are highlighted with a red box, while bystander positions are indicated with black boxes. **C.** CD39: relative β-globin mRNA levels as evaluated by RT-qPCR in erythroid cells derived from corrected β-thalassemic HSPCs (cor). As controls, we used erythroid cells derived from β-thalassemic or HD HSPCs transfected only with NRCH/SpRY-ABE8e mRNA (BE) (3 β-thalassemia patients and 3 HDs). IVS2-1: relative β-globin mRNA levels as evaluated by RT-qPCR using primers amplifying exclusively correctly spliced β-globin mRNAs in erythroid cells derived from corrected β-thalassemic HSPCs (cor). As controls, we used erythroid cells derived from β-thalassemic or HD HSPCs transfected only with SpRY-ABE8e mRNA (BE) (1 β-thalassemia patients and 3 HDs). β-globin expression was normalized to α-globin. Data are expressed as mean±SEM. Dotted lines indicate maximum and minimum values observed in HD cells. **D.** RT-PCR using primers amplifying a region from *HBB* exon 1 to exon 3. cDNA was obtained from erythroid cells derived from corrected β-thalassemic HSPCs (cor). As controls, we used erythroid cells derived from β-thalassemic or HD HSPCs transfected only with SpRY-ABE8e mRNA (BE) (n=2 biologically independent experiments, 1 β-thalassemia patient and 3 HDs). **E.** Expression of β-, β-Nîmes, δ-, γG- and γA- globin chains measured by RP-HPLC in β-thalassemic and HD RBCs. β-like-globin expression was normalized to α-globin. The α-/non-α-globin ratio is reported on top of the graph. RBCs were obtained from corrected β-thalassemic HSPCs (cor). As controls, we used RBCs derived from β-thalassemic patients' or HD HSPCs transfected only with SpRY/NRCH-ABE8e mRNA (BE) (CD39: n=2 biologically independent experiments, 3 donors, D13 of the differentiation; IVS2-1: n=2 biologically independent experiments, 1 donor, BT3.a= D20, BT3.b= D16 of the differentiation). Data are expressed as mean±SEM. **F.** Analysis of HbA, Hb-Nîmes, HbF and HbA_2 by CE-HPLC in β-thalassemic patients and HD RBCs. We calculated the percentage of each Hb type over the total Hb tetramers. RBCs were obtained from corrected β-thalassemic HSPCs (cor). As controls, we used RBCs derived from β-thalassemic patients' or HD HSPCs transfected only with SpRY/NRCH-ABE8e mRNA (BE) (CD39: n=2 biologically independent experiments, 3 donors, D16 of the differentiation; IVS2-1: n=2 biologically independent experiments, 1 donor, BT3.a=

D20, BT3.b= D16 of the differentiation). HD data are expressed as mean±SEM. BT, β-thalassemia patients. HD, healthy donors.

Figure 3. Efficient reversion of the CD39 (CAG>TAG) and IVS2-1 (G>A) mutations in β-thalassemic HSPCs corrects ineffective erythropoiesis.

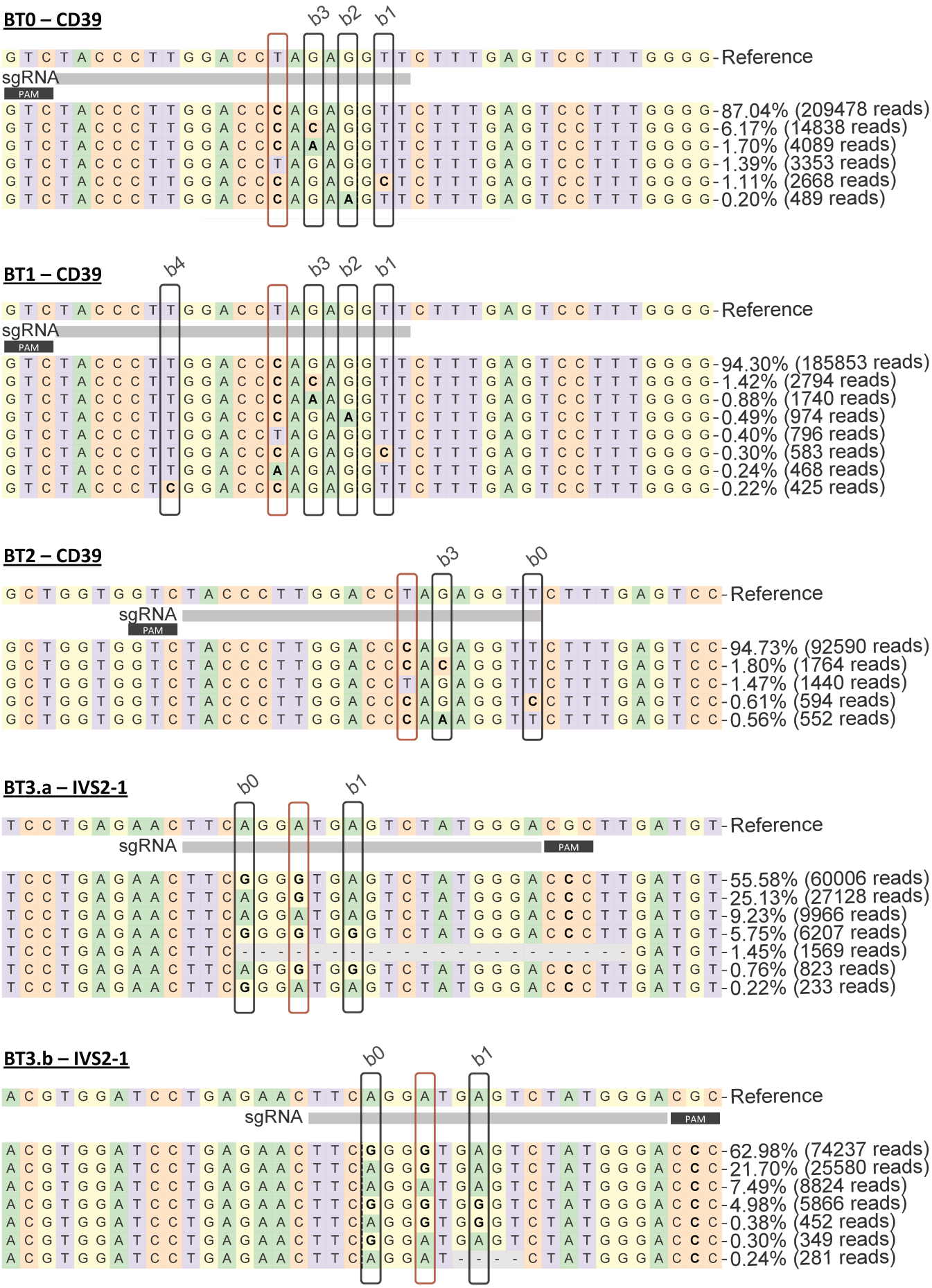
A. Frequency of GPA^+, $CD71^+$ and $CD36^+$ cells at day 19 of erythroid differentiation, as measured by flow cytometry analysis. As controls, we used RBCs derived from β-thalassemic patients' or HD HSPCs transfected with TE only (TE) or with SpRY/NRCH-ABE8e mRNA only (BE) (CD39: n=2 biologically independent experiments, 3 HDs and 3 BT donors; IVS2-1: n=2 biologically independent experiments, 3 HDs and 1 BT donor). Data are expressed as mean±SEM **B.** Frequency of enucleated cells at day 13, 16 and 19 of erythroid differentiation, as measured by flow cytometry analysis of cells stained with the DRAQ5 nuclear dye. As controls, we pooled data obtained in RBCs derived from β-thalassemic patients' or HD HSPCs transfected with TE or only with SpRY-ABE8e/NRCH-ABE8e mRNA (ctr) (CD39: n=2 biologically independent experiments, 3 HDs and 3 BT donors; IVS2-1: n=2 biologically independent experiments, 3 HDs and 1 BT donor). Data for HD samples are expressed as mean±SEM. **C.** Cell size of enucleated cells at day 16 and 19 of erythroid differentiation, as measured by flow cytometry using the median of FSC-A intensity. As controls, we used RBCs derived from β-thalassemic patients' or HD HSPCs transfected with TE or only with SpRY-ABE8e/NRCH-ABE8e mRNA (ctr) (CD39: n=2 biologically independent experiments, 3 HDs and 3 BT donors; IVS2-1: n=2 biologically independent experiments, 3 HDs and 1 BT donor). Data for HD samples are expressed as mean±SEM. **D.** MFI in ROS-containing cells ($DCFDA^+$) for control (BT-ctr) and edited (BT-cor) β-thalassemic samples. For the controls (ctr), we pooled data obtained in RBCs derived from β-thalassemic HSPCs transfected with TE or with SpRY/NRCH-ABE8e mRNA only (CD39: n=2 biologically independent experiments, 3 donors; IVS2-1: n=2 biologically independent experiments, 1 BT donor). * $p \leq 0.05$ (Paired T-test). **E.** Flow cytometry histograms showing the frequency of apoptotic cells ($AnnexinV^+$-cells) in the $7AAD^-$ cell population in unstained (Uns), β-thalassemic corrected (BT-cor) or treated only with TE (BT-TE) at day 13 of erythroid differentiation. As controls, we used RBCs derived from HD HSPCs transfected with TE only (HD-TE) (CD39: n=2 biologically

independent experiments, 3 HDs and 3 BT donors; IVS2-1: n=2 biologically independent experiments, 3 HDs and 1 BT donor).

F. Supplemental files

1. Supplemental figures

Supplemental Figure 1



2. Supplemental figure legends.

Supplementary figure 1. On-target editing in β-thalassemic cells.

Frequency and sequence of modified and unmodified alleles in corrected β-thalassemic samples, as measured by targeted NGS sequencing. Target base position is highlighted with a red box. Bystander edits are highlighted with black boxes.

G. Materials and Methods

HSPC and T cell purification and culture

We obtained human non-mobilized peripheral blood $CD34^+$ HSPCs from β-thalassemia patients. Samples eligible for research purposes were obtained from the "Hôpital Necker-Enfants malades" Hospital (Paris, France). Written informed consent was obtained from all adult subjects. All experiments were performed in accordance with the Declaration of Helsinki. The study was approved by the regional investigational review board (reference: DC 2014-2272, CPP Ile-de-France II "Hôpital Necker-Enfants malades"). HSPCs were purified by immunomagnetic selection immunostaining with the CD34 MicroBead Kit (Miltenyi Biotec). The $CD34^-$ fraction was kept for T cell cultures. Forty-eight hours before transfection, $CD34^+$ cells were thawed and cultured at a concentration of $5x10^5$ cells/mL in the "HSPC medium" containing StemSpan (STEMCELL Technologies) supplemented with penicillin/streptomycin (Gibco), 250 nM StemRegenin1 (STEMCELL Technologies), and the following recombinant human cytokines (PeproTech): human stem cell factor (SCF) (300 ng/ml), Flt-3L (300 ng/ml), thrombopoietin (TPO) (100 ng/ml), and interleukin-3 (IL-3) (60 ng/ml). Four days before transfection, the $CD34^-$ fraction was thawed and cultured at $5x10^6$ cells/mL in the "T cells medium" containing RPMI 1640 + GlutaMAX (Gibco) supplemented with FBS (Thermo), penicillin/streptomycin (Gibco) and Recombinant Human IL-2 (Peprotech). After recovery, cells were transferred to "T cell activation medium" supplemented with CD28 Monoclonal Antibody (eBioscience, Clone CD28.2) in plates coated with CD3 Monoclonal Antibody (eBioscience, Clone OKT3).

Base editor plasmids

Constructs used in this study include NG-ABE8e (Plasmid #138491, Addgene), NRCH-ABE8e and SpRY-ABE8e plasmids. The NRCH-ABE8e plasmid was created by replacing the Cas9-coding sequence of the ABE8e plasmid (Plasmid #138489, Addgene)[82] plasmid with the Cas9-NRCH included in the "pCMV-ABEmax-NRCH" plasmid (Plasmid #136923, Addgene)[80]. The SpRY-ABE8e plasmid was created by replacing the Cas9-coding sequence of the ABE8e plasmid (Plasmid #138489,

Addgene)[82] with the Cas9 fused to GFP included in the "pCMV-T7-ABEmax(7.10)-SpRY-P2A-EGFP (RTW5025)" plasmid (Plasmid #140003, Addgene)[81].

gRNA design

We manually designed gRNAs targeting the CD39 and IVS2-1 mutations (**Table 1**). We used chemically modified synthetic gRNAs harboring 2′-O-methyl analogs and 3′-phosphorothioate nonhydrolyzable linkages at the first three 5′ and 3′ nucleotides (Synthego).

Table 1. gRNA target sequences.

gRNA	Sequence (5' to 3')	PAM sequence (5' to 3')	Position (hg19)	Strand	Compatible ABEs
CD39 gRNA1	AACCTCTAGGTCCAAGGGTA	GACC	chr11 : 5247997 – 5248016 (*HBB*)	+	NRCH-ABE8e, SpRY-ABE8e
CD39 gRNA2	ACCTCTAGGTCCAAGGGTAG	ACCA	chr11 : 5247998 – 5248017 (*HBB*)	+	SpRY-ABE8e
CD39 gRNA3	CCTCTAGGTCCAAGGGTAGA	CCAC	chr11 : 5247999 – 5248018 (*HBB*)	+	SpRY-ABE8e
CD39 gRNA4	CTCTAGGTCCAAGGGTAGAC	CACC	chr11 : 5248000 – 5248019 (*HBB*)	+	NRCH-ABE8e, SpRY-ABE8e
CD39 gRNA5	TCTAGGTCCAAGGGTAGACC	ACCA	chr11 : 5248001 – 5248020 (*HBB*)	+	SpRY-ABE8e
IVS2-1 gRNA1	TCAGGATGAGTCTATGGGAC	GCTT	chr11 : 5247792 – 5247811 (*HBB*)	-	SpRY-ABE8e
IVS2-1 gRNA2	TTCAGGATGAGTCTATGGGA	CGCT	chr11 : 5247793 – 5247812 (*HBB*)	-	NG-ABE8e, SpRY-ABE8e
IVS2-1 gRNA3	CTTCAGGATGAGTCTATGGG	ACGC	chr11 : 5247794 – 5247813 (*HBB*)	-	SpRY-ABE8e

mRNA in vitro transcription

20 µg of NG-ABE8e, NRCH-ABE8e or SpRY-ABE8e expressing plasmids were digested overnight with SapI restriction enzyme (Thermo) that cleaves once right after the poly-A tail. The linearized plasmids were purified using a PCR purification kit (QIAGEN #28106) and were eluted in 14 µl of DNase/RNase-free water. 2 µg of linearized plasmid were used as template for the *in vitro* transcription reaction (MEGAscript, Ambion #AM1334). The *in vitro* transcription protocol was modified as follows. The GTP nucleotide solution was used at a final concentration of 3.0 mM instead of 7.5 mM and the anti-reverse cap analog N7-Methyl-3'-O-Methyl-Guanosine-5'-Triphosphate-5'-Guanosine (ARCA, Trilink #N-7003) was used at a final concentration of 12.0 mM resulting in a final ratio of Cap:GTP of 4:1 that allows efficient mRNA capping. The incubation time for the *in vitro* reaction was reduced to

30 minutes. mRNA was precipitated using lithium chloride and resuspended in TE buffer in a final volume that allowed to achieve a concentration of >1 µg/µl. The mRNA quality was assessed using Bioanalyzer (Agilent).

RNA transfection

$1x10^6$ T cells per condition were transfected with 3.0 µg of the ABE-encoding mRNA and 3.2 µg of the synthetic gRNA. When ABE was not fused to GFP, a GFP-encoding mRNA (Tebu-bio) was added to the transfection mix. We used the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the EO115 program (Nucleofector 4D). Cells transfected only with TE buffer served as negative controls.

$1x10^4$ to $5x10^5$ HSPCs per condition were transfected with 3.0 µg of the ABE-encoding mRNA and 3.2 µg of the synthetic gRNA. We used the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the CA137 program (Nucleofector 4D). Cells transfected only with TE buffer or only with the ABE-encoding mRNA served as negative controls.

HSPC differentiation

Transfected $CD34^+$ HSPCs were differentiated into mature red blood cells (RBCs) using a three-phase erythroid differentiation protocol, as previously described[156,179]. During the first phase (day 0 to day 6), cells were cultured in a basal erythroid medium supplemented with 100 ng/ml recombinant human SCF (PeproTech), 5 ng/ml recombinant human IL-3 (PeproTech), 3 IU/ml EPO Eprex (Janssen-Cilag) and 10^{-6} M hydrocortisone (Sigma). During the second phase (day 6 to day 9), cells were co-cultured with MS-5 stromal cells in the basal erythroid medium supplemented with 3 IU/ml EPO Eprex (Janssen-Cilag). During the third phase (day 9 to day 20), cells were co-cultured with stromal MS-5 cells in a basal erythroid medium without cytokines. Erythroid differentiation was monitored by flow cytometry analysis of CD36, CD71, GYPA and of enucleated cells using the DRAQ5 double-stranded DNA dye. 7AAD was used to identify live cells.

Evaluation of editing efficiency

Base editing efficiency and InDel frequency were evaluated in HSPC-derived erythroid cells at D13 of differentiation, and in T cells 3 days after transfection.

Genomic DNA was extracted from control and edited cells using PURE LINK Genomic DNA Mini kit (LifeTechnologies), or Quick-DNA/RNA Miniprep (ZYMO Research), following manufacturers' instructions. To evaluate base editing efficiency in T cells, we performed either a nested PCR using previously published primers for CD39[170], or a regular PCR with newly designed primers for IVS2-1 (**Table 2**) followed by Sanger sequencing and EditR analysis[180]. TIDE analysis (Tracking of InDels by Decomposition) was performed to evaluate the percentage of InDels in edited samples[181].

On-target regions of DNA from HSPC-derived erythroid cells were PCR-amplified and subjected to NGS. On-target sites were PCR-amplified using the Phusion High-Fidelity polymerase (NEB, M0530) and primers containing specific DNA stretches (MR3 for forward primers and MR4 for reverse primers; **Table 2**) located 5' to the sequence recognizing the off-target. For the on-target site of CD39, a nested PCR was applied. Amplicons were purified using Ampure XP beads (Beckman Coulter, A63881). Illumina-compatible barcoded DNA amplicon libraries were prepared by a second PCR step using the Phusion High-Fidelity polymerase (NEB, M0530) and primers containing Unique Dual Index (UDI) barcodes and annealing to MR3 and MR4 sequences. Libraries were pooled, purified using the High Pure PCR Product Purification Kit (Sigma-Aldrich, 11732676001), and sequenced using Illumina NovaSeq 6000 system (paired-end sequencing; 2×100-bp) to obtain a minimum of 100,000 reads per amplicon. Targeted NGS data were analyzed using CRISPResso2[210].

Table 2. Primers used for on-target sequencing.

Amplified region	F/R	Sequence (5' to 3')
IVS2-1	F	GGATCTGTCCACTCCTGATGC
	R	ACTGTACCCTGTTACTTATCCCC
NGS_IVS2-1	F	GCAGCGTCAGATGTGTATAAGAGACAGGTACCCTGTTACTTATCCCCTTCC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTCAAGGGCACCTTTGCCAC
NGS_CD39-1	F	CTCCTGAGGAGAAGTCTGCCGTTAC
	R	GCAGCTCACTCAGTGTGGC
NGS_CD39-2	F	GCAGCGTCAGATGTGTATAAGAGACAGGACTCTTGGGTTTCTGATAGGCAC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGAGCCTTCACCTTAGGGTTGCC

F, forward primer; R, reverse primer.

RT-PCR and RT-qPCR

RNA was extracted from cells at day 13 of differentiation (Qiagen, 74004 or Zymo Research, ZD7001) and retro-transcribed (Thermo, 18080051). RT-PCR analysis of β-globin mRNAs was performed using previously described primers spanning exon 1 to 3[170]. RT-qPCR was performed using previously described primers for the detection of correctly spliced β-globin mRNA[211] and the following primers amplifying α-globin transcripts: α-globin-F 5'-CGGTCAACTTCAAGCTCCTAA-3' and α-globin-R 5'-ACAGAAGCCAGGAACTTGTC-3'.

Flow cytometry analysis

Flow cytometry analysis of CD36, CD71 and GYPA erythroid surface markers on HSPC-derived erythroid cells was performed using a V450-conjugated anti-CD36 antibody (561535, BD Horizon), a FITC-conjugated anti-CD71 antibody (555536, BD Pharmingen) and a PE-Cy7-conjugated anti-GYPA antibody (563666, BD Pharmingen). Flow cytometry analysis of enucleated or viable cells was performed using double-stranded DNA dyes (DRAQ5, 65-0880-96, Invitrogen and 7AAD, 559925, BD, respectively). Apoptosis was evaluated using PE Annexin V Apoptosis Detection Kit I (BD Biosciences). Flow cytometry analyses were performed using Gallios (Beckman coulter) flow cytometer. Data were analyzed using the FlowJo (BD Biosciences) software.

RP-HPLC analysis of globin chains

Reversed-phase HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). A 250x4.6 mm, 3.6 µm Aeris Widepore column (Phenomenex) was used to separate globin chains by HPLC. Samples were eluted with a gradient mixture of solution A (water/acetonitrile/trifluoroacetic acid, 95:5:0.1) and solution B (water/acetonitrile/trifluoroacetic acid, 5:95:0.1). The absorbance was measured at 220 nm.

CE-HPLC analysis of hemoglobin tetramers

Cation-exchange HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). A 2 cation-exchange column (PolyCAT A, PolyLC, Columbia, MD) was used to separate hemoglobin

tetramers by HPLC. Samples were eluted with a gradient mixture of solution A (20mM bis Tris, 2mM KCN, pH=6.5) and solution B (20mM bis Tris, 2mM KCN, 250mM NaCl, pH=6.8). The absorbance was measured at 415 nm.

V. Development of a base editing strategy to correct the SCD mutation.

A. Introduction

With more than 300,000 affected children born every year, SCD is the most prevalent monogenic disorder around the world[212]. It is caused by a point mutation in the *HBB* gene, which leads to the production of the defective β_S-globin. The resulting sickle hemoglobin tetramers (HbS) polymerize under hypoxia, and RBCs assume a sickle shape and become inflexible. Sickle RBCs have a short lifespan and obstruct micro vessels causing a chronic multi-organ disease associated with poor quality of life and short life expectancy. Transplantation of autologous, genetically modified Hematopoietic Stem and Progenitor Cells (HSPCs) was investigated as a treatment option[2]. However, current gene therapy strategies based on the use of lentiviral vectors or CRISPR/Cas9 nuclease are not equally effective in all the patients and/or raise safety concerns[213]. In particular, vector integration can cause insertional mutagenesis and CRISPR/Cas9 mediated DSB generation can induce cell toxicity and genomic rearrangements[61].

In SCD, the causative mutation replaces the 6th codon of the *HBB* gene GAG encoding a glutamic acid with a GTG codon encoding a valine. As ABEs can convert A●T base pairs to G●C, they allow the conversion of the valine-specifying GTG codon to an alanine-specifying GCG codon. Although this conversion does not restore the original amino acid, it does allow the introduction of an amino acid found in a known variant of HbA called HbG-Makassar[136–139]. This variant is presumed to be functional as it was described in homozygous healthy individuals[138].

However, at the time of the initiation of this study, a major hurdle to overcome was the inaccessibility of the SCD mutation to canonical ABEs. In fact, with no canonical PAM sequence being located close to the target base, we aimed to engineer an ABE in order to access the target base. Similarly to others, our work culminated in an efficient ABE able to achieve high mutation correction efficiencies in SCD HSPCs, raising the possibility to develop an efficient gene correction treatment for SCD[127,214].

B. Results

HbG-Makassar cannot polymerize

First, we investigated how the alanine in β-Mak could avoid Hb polymerization. HbS polymerization is due to the replacement of the hydrophilic glutamic acid by the hydrophobic valine in position 6 (V6), which then interacts with the hydrophobic pocket formed by 3 amino acids of an adjacent β^s-globin (**Figure 1A**; **Figure 1B**, black and red panels). The alanine in position 6 (A6) of β-Mak globin is also a hydrophobic amino acid (**Figure 1A**). However, structural predictions indicated that the alanine of the β-Mak is not bulky enough to fit in this pocket and thus the non-physiological interaction between hemoglobins does not occur (**Figure 1B**, blue panel).

Negligible gene correction efficiency using canonical ABEs

Base editors available at the time of the initiation of this study consisted in the original ABE (ABE7.10[74]), which was optimized for expression and nuclear localization (ABEmax[161]), but also further engineered to recognize the non-canonical NG PAM (xABEmax and NG-ABEmax containing the xCas9 and NG-Cas9, respectively[77]). We identified two compatible PAMs near to the SCD mutation and two gRNAs were designed accordingly: HBS_A and HBS_B (**Figure 2A, Table 1**). However, while these gRNAs included the SCD mutation within their sequence, the mutation was not located within the optimal editing window corresponding to positions 4-8 of the protospacer (**Figure 2B**). A first screening was performed in an erythroid cell line carrying the SCD mutation (SCD-HUDEP2) with compatible enzymes using plasmid delivery (**Figure 2B**). Even though some A>G conversions were observed within the editing windows, the level of editing did not exceed background noise at the target base located upstream (HBS_B) or downstream (HBS_A) of the editing window (**Figure 2B**).

Extending gRNA length increases editing efficiency

Previous reports showed that adding a few nucleotides at the 5' end of the gRNA sequence could increase base editing efficiency in the first positions of the protospacers probably by improving deaminase accessibility (**Figure 2C**)[215,216]. This strategy was tested with gRNA HBS_B that places the SCD mutation in position 2 in

SCD-HUDEP2. By extending HBS_B from 20 to 21 and 22 nucleotides, we observed no increase in efficiency at the target base with none of the tested enzymes (ABEmax, xABEmax or NG-ABEmax) (**Figure 2D**). Increasing further the length of the gRNA up to 29 nucleotides led to an overall decrease in efficiency when using ABEmax (**Figure 2D**). However, it is noteworthy that with all the three enzymes a significant increase in the efficiency was observed at position 4 with the 21-nucleotide-long gRNA and to a lower extent with the 22-nucleotide-long gRNA, compared to the original 20-nucleotide-long gRNA. On the contrary, the efficiency of extended gRNA at position 7 was decreased compared to the original 20-nucleotide-long gRNA. Overall, these data show that increasing the length of the gRNA can be used as a tool to modulate the efficiency within the editing window, and in particular to increase base editing efficiency in position 4.

Increasing the length of the linker sequence improves editing efficiency

Different teams observed that modifying the linker region located between the Cas9n and the deaminase in CBEs could increase editing efficiency and shift the editing window in the region close to the PAM (**Figure 2E**)[83,84,217]. In particular, a substantial increase in base editing efficiency was obtained by inserting the single-strand DNA binding domain (ssDBD) Rad51 or two extra nuclear localization signals (NLS) (2X) in the original CBE linker sequence[83,84]. Therefore, we decided to apply these modifications to ABEs. Given the expected increase in base editing efficiencies toward the PAM, we first implemented these linker modifications to NG-ABEmax, an enzyme compatible with HBS_A, which places the target base (the SCD mutation) downstream of the editing window, closer to the PAM (position 9) (**Figure 2B**). While editing efficiencies with the regular NG-ABEmax enzyme were not exceeding background noise, we observed an improvement of editing efficiencies at the target base with NG-ABEmax-2X, but not with NG-ABEmax-Rad51 in SCD-HUDEP2 cells (**Figure 2F**). We then tested NG-ABEmax-2X in parallel with gRNA HBS_B (which places the target base/SCD mutation upstream of the editing window) and observed an overall increase in base editing efficiencies within the editing window but also at the target base upstream of the editing window, allowing gene correction (**Figure 2G**). Overall, these results demonstrate that addition of the 2X linker in ABEs, could be used to increase base editing efficiencies and modulate the editing window. In this way, the SCD mutation was corrected with an efficiency of ~15% using HBS_A and

HBS_B, two gRNAs placing the SCD mutation out of their original base editing window.

The use of the PAM-less Cas9 allowed the design of efficient gRNAs

During our study, Walton et al. developed a near PAM-less Cas9 (SpRY), recognizing all NRN PAMs, and also, to a lower extent NYN PAMs[81]. This novelty allowed the unconstrained design of 5 gRNAs placing the SCD mutation in positions 4-8 of the protospacer, i.e. within the editing window (**Figure 3A, Table 1**). SpRY-ABEmax was used to screen these gRNAs in SCD-HUDEP2 (**Figure 3A**). We achieved good gene correction efficiencies, i.e. 30% and 16% with HBS_F and HBS_G, respectively (**Figure 3B**).

The use of a highly processive deaminase substantially increased gene correction efficiency

Recently, Richter et al. developed the highly processive enzyme ABE8e by inserting 8 mutations within the TadA domain of the deaminase[82]. First, we used NG-ABE8e in which the original deaminase of NG-ABEmax was replaced with this optimized deaminase[82] using HBS_A and HBS_B in SCD-HUDEP2, which place the target base outside the editing window (**Figure 3C**). Modification of the deaminase allowed efficient editing at the target base (49% and 32% with HBS_A and HBS_B, respectively), even though if it is placed outside of the original editing window. Moreover, a substantial increase in base editing efficiencies within the editing window was also observed, leading to the generation of bystander edits whose potential consequences are discussed below (see **Bystander editing**) (**Figure 3C**).

PAM-less Cas9, highly processive deaminase and extended gRNA synergize in increasing gene correction efficiency

After identifying critical parameters allowing the efficient improvement of A>G conversions at the target base, we aimed to combine them in order to reach higher efficiencies. First, we combined the highly processive deaminase ABE8e with the near PAM-less Cas9 SpRY generating a new construct named ABE8e-SpRY. As previously observed (**Figure 3C**)[82], modification of the deaminase led to a substantial increase in base editing efficiencies with gRNAs HBS_C to G in SCD-HUDEP2 (**Figure 3D**). The 3 most efficient gRNAs were HBS_C, HBS_F and HBS_G, with a

gene correction frequency of 85, 88 and 89%, respectively. Next, as the gRNA HBS_C places the target mutation at position 4, we thought to further increase the correction efficacy at this position by using a 21-nucleotide-long version of HBS_C, hereinafter referred to as HBS_C_21. Interestingly, extension of the gRNA, as previously observed, led to a significant increase in the editing rate of position 4 in SCD-HUDEP2, reaching 92% (**Figure 3E**). We then tested the three most efficient combinations (ABE8e-SpRY with HBS_C_21, HBS_F and HBS_G) in SCD HSPCs using RNA delivery. We achieved a good base editing efficiency, although it did not exceed 80% and was lower than in SCD-HUDEP2 cells (**Figure 3F**).

Extending the linker as a potential strategy to improve editing efficiency

As we previously observed (**Figure 2G**) that the 2X linker addition could increase base editing efficiencies within the editing window, we added the 2X linker within the ABE8e-SpRY enzyme in order to evaluate if we could further increase efficiencies in SCD-HSPCs (**Figure 3G**). We tested this new construct in SCD-HSPCs in combination with the three most potent gRNAs (HBS_F, G and C_21) (**Figure 3H**). Preliminary data showed that 2X linker addition to ABE8e-SpRY led to a decrease in efficiencies from positions 4 to 8 of the protospacer (corresponding to the original editing window), while leading to an increase from positions 9 to 13. Thus, we concluded that the 2X linker addition to ABE8e-SpRY was not an efficient way to further increase the mutation correction rate of this strategy. In addition, it could lead to higher conversion rates of adjacent non-targeted adenines (bystander edits, see paragraph "**Bystander editing**").

Bystander editing

While improving editing efficiencies at the on-target base (n), we also observed significant base conversion in three adjacent adenines in positions n+2, n+5 and n+9 (**Figure 3A**). While both n+2 and n+5 A>G conversions create silent mutations, the n+9 edit is responsible for a missense mutation replacing a leucine with a proline. This single amino acid change is found in a Hb variant known as Hb Jabalpur[218]. Importantly, this variant was previously reported in a heterozygous individual with a mild anemia and a low ratio of Hb Jabalpur over total Hbs (~30%), suggesting that this mutation confers a disadvantage (e.g., instability) to the Hb[219]. However, current strategies demonstrate a low editing rate at this position (<10%, **Figure 3F)**.

C. Discussion

Given its monogenic nature and high prevalence, development of gene therapy approaches aiming to correct the SCD mutation benefits from a very dynamic scientific context. In the past decades, several HDR-based strategies were developed in order to correct the unique causative mutation of SCD[130,131,220,221]. While these strategies were limited by poor HDR efficiency in HSCs, base editing demonstrated higher editing rates and maintenance in long-term repopulating HSCs[126–128]. In this work, we aimed to develop an efficient base editing strategy to correct the SCD mutation by generating the non-pathogenic HbG-Makassar variant. As this mutation is difficult to access with first generation base editors given the absence of proximal canonical PAMs, we sequentially engineered different domains of ABE enzymes in order to modulate their activity and we gradually increased the editing efficiency at the target base position. Interestingly, this study reflects the complexity of the design and optimization of the base editing process. In fact, none of the optimized parameters by itself allowed to exceed 30% of gene correction. However, the combination of different parameters has resulted in correction efficiencies of around ~90%. Therefore, our work gathers some of the most recent optimizations of the base editing tools and demonstrate how they can be combined in order to efficiently reach difficult-to-edit positions.

While carrying out this study, two papers came out demonstrating correction of the SCD mutation in SCD HSPCs. The first study, from Chu et al., described the rational design of an ABE variant able to shift the editing window and reach the SCD mutation while using gRNA HBS_A compatible with an NG PAM[214]. In addition, Newby et al. used HBS_F in combination with NRCH-ABE8e, an enzyme combining the highly processive deaminase ABE8e and the evolved Cas9 recognizing NRCH PAM[80,127]. Both studies showed gene correction similar to ours (~80%) in SCD HSPCs.

Future works will aim to further characterize our strategy in terms of correction of the SCD phenotype in SCD HSPC-derived erythroid cells, persistence of gene correction events in long-term repopulating HSCs and off-target activity. Importantly, published works already provide some insights as they observed that, with similar levels of

HbG-Makassar allele conversion, the SCD phenotype was corrected in erythroid cells obtained from *in vitro* differentiated edited SCD-HSPCs[127,214]. It would be interesting to compare the Makassar-inducing strategies with other existing strategies - such as other base editing strategies aiming to reactivate HbF[126] (**Annex 1**) -, in terms of correction of the phenotype and editing efficiency in long-term repopulating HSCs.

As our enzyme relies on a near PAM-less Cas9 for target recognition, additional off-targets might be detected compared to the two other developed strategies based on enzymes recognizing a more restrictive PAM. Newby et al. detected high gRNA-dependent off-target editing activity with no anticipated consequences[127]. Preliminary *in silico* off-target nomination using COSMID[182], identified a potentially reduced off-target activity with HBS_C_21 due to its longer sequence. However, the effect of the additional nucleotide on off-target activity is difficult to predict and will need to be confirmed experimentally. In fact, a previous study demonstrated that, despite fewer unique on-target sites, truncated gRNAs showed an improved specificity[222].

While Chu et al. do not provide an extended off-target analysis for this strategy, they observed that their construct exhibits less gRNA-independent DNA and RNA deamination as the structural modifications implemented to the enzyme constrained the deaminase activity to focus to the R-loop-generated-ssDBD[214]. Similarly, off-target activity of our enzyme could be mitigated by the 2X linker usage. In fact, the addition of a ssDBD linker in a ABE8e derivative developed for base editing in plants showed reduced off-target activity[223]. Thus, the stronger affinity of the construct for the targeted DNA may influence the spatial structure of the DNA-bound complex and improve the specificity by restricting the deamination capacity of the TadA domain to strictly matched target sequences. While the 2X linker (containing two NLS) was originally introduced to increase nuclear localization of the base editor, this effect was not observed[84]. However, the positively charged NLS sequences could facilitate the interaction with the negatively charged DNA backbone, thus this increased affinity could also mitigate off-target activity. While 2X linker addition to ABE8e-SpRY sensibly decreased on-target activity using HBS_C_21 in SCD-HSPCs, this was poorly affected with HBS_F and G and did not increase the bystander edit leading to a missense mutation (Leu>Pro) (**Figure 3H**). Thus, this strategy, if validated in mammalian cells, could be implemented with HBS_F or G gRNAs. Finally, if off-target

assessment reveals potentially deleterious off-target effects, further engineering of the enzyme could be performed to include mutations known to increase Cas9 specificity[93,94].

Bystander edits are a challenge when aiming to precisely correct a sequence with base editing. In particular, in this and other works, we identified A>G conversions occurring close to the on-target base editing. Among these, one leads to a non-synonymous mutation that confers a disadvantage to the resulting Hb[219]. Thus, selection of a clinically relevant strategy will need to take into consideration the occurrence of this bystander edit as a potential factor limiting the efficacy of the approaches generating the HbG-Makassar.

Lastly, our future work aims to provide further insights into the structure and functionality of the HbG-Makassar. Bioinformatic predictions described in this work anticipate a structural resistance of this Hb variant to sickling. However, before clinical application, a deeper study of the native HbG-Makassar protein is required. In particular, a complete description of biochemical and biophysical properties will further support the clinical applicability of this gene correction strategy. Moreover, a potential concern of this strategy is the potential immunogenicity of the HbG-Makassar protein. However, the long-term experience of chronic RBC transfusion in SCD patients has demonstrated the relatively low immunogenic potential of Hb, aside from the description of the development of a humoral response against HbA in 3 patients suggesting that a single amino-acid change can be immunogenic[224]. Importantly, this study did not indicate that the presence of the antibody has contributed to the severity of SCD, or to a shortening of survival of transfused erythrocytes[224]. Of note, “prime editing” is a new genome editing method, which allows the insertion of precise insertions, deletions and all types of point mutations without requiring DSBs or donor DNA templates [225]. Thus, it has been used to precisely and efficiently correct the SCD mutation in cell lines[225]. While this strategy would correct the *HBB* gene to its WT form, it is poorly efficient in primary cells so far[226], and would require further optimizations to reach therapeutically relevant corrections rates.

D. Figures

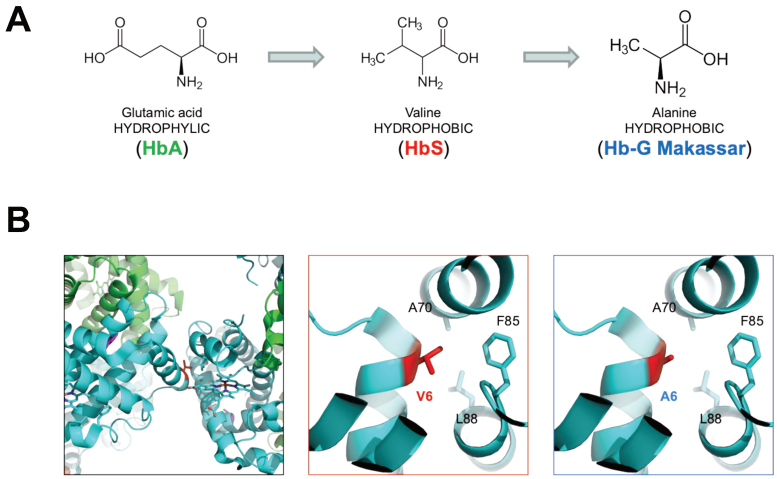

A
Glutamic acid
HYDROPHYLIC
(HbA)
Valine
HYDROPHOBIC
(HbS)
Alanine
HYDROPHOBIC
(Hb-G Makassar)
B
A70
F85
V6
L88
A70
F85
A6
L88

Figure 1

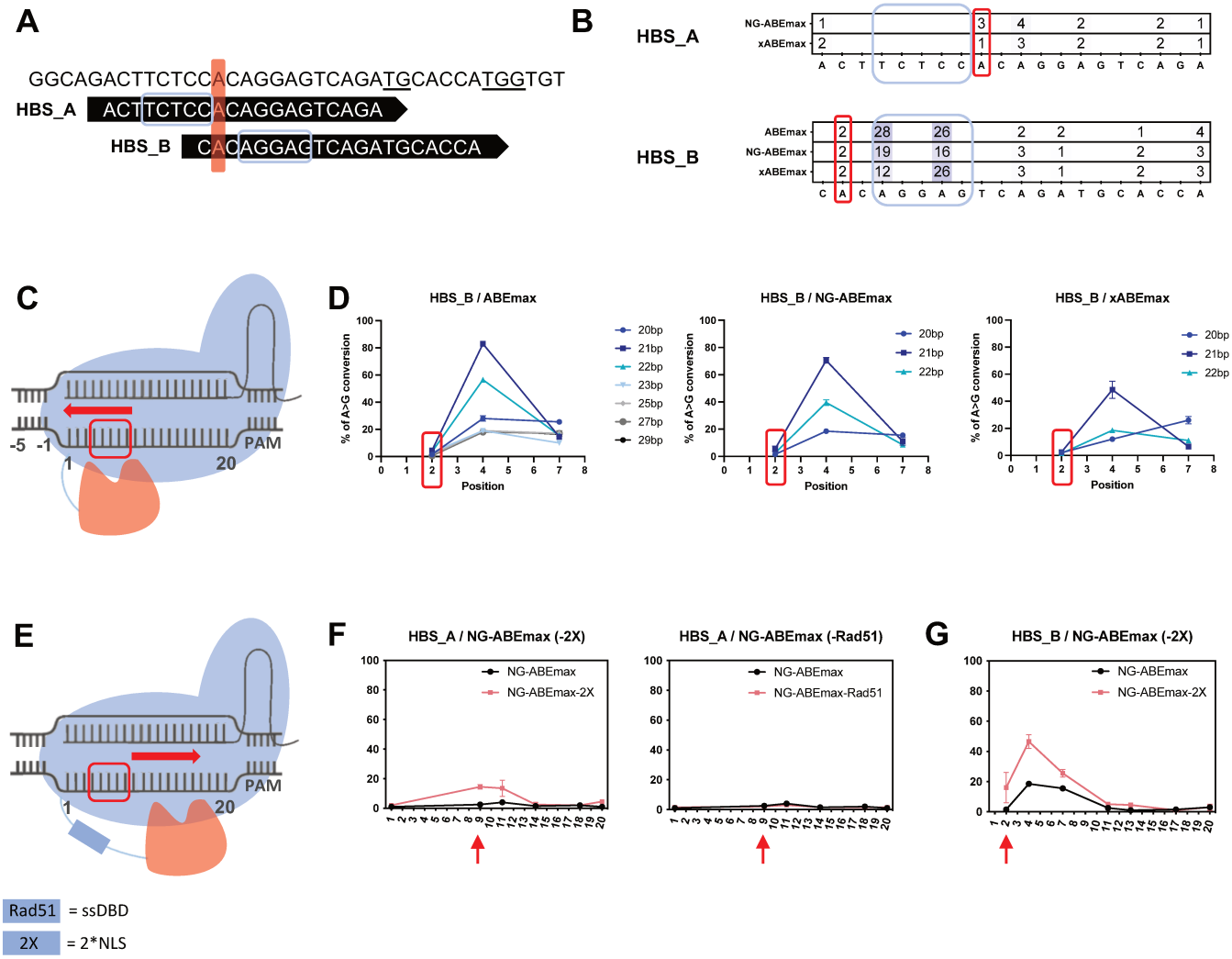

A
GGCAGACTTCTCCACAGGAGTCAGATGCACCATGGTGT
HBS_A ACTTCTCCACAGGAGTCAGA
HBS_B CACAGGACTCAGATGCACCA
B
HBS_A
NG-ABEmax 1 3 4 2 2 1
xABEmax 2 1 3 2 2 1
A C T T C T C C A C A G G A G T C A G A
HBS_B
ABEmax 2 28 26 2 2 1 4
NG-ABEmax 2 19 16 3 1 2 3
xABEmax 2 12 26 3 1 2 3
C A C A G G A G T C A G A T G C A C C A
C
-5 -1 PAM
1 20
D
HBS_B / ABEmax
HBS_B / NG-ABEmax
HBS_B / xABEmax
% of A>G conversion
Position
20bp
21bp
22bp
23bp
25bp
27bp
29bp
E
PAM
1 20
F
HBS_A / NG-ABEmax (-2X)
HBS_A / NG-ABEmax (-Rad51)
NG-ABEmax
NG-ABEmax-2X
NG-ABEmax-Rad51
G
HBS_B / NG-ABEmax (-2X)
Rad51 = ssDBD
2X = 2*NLS

Figure 2

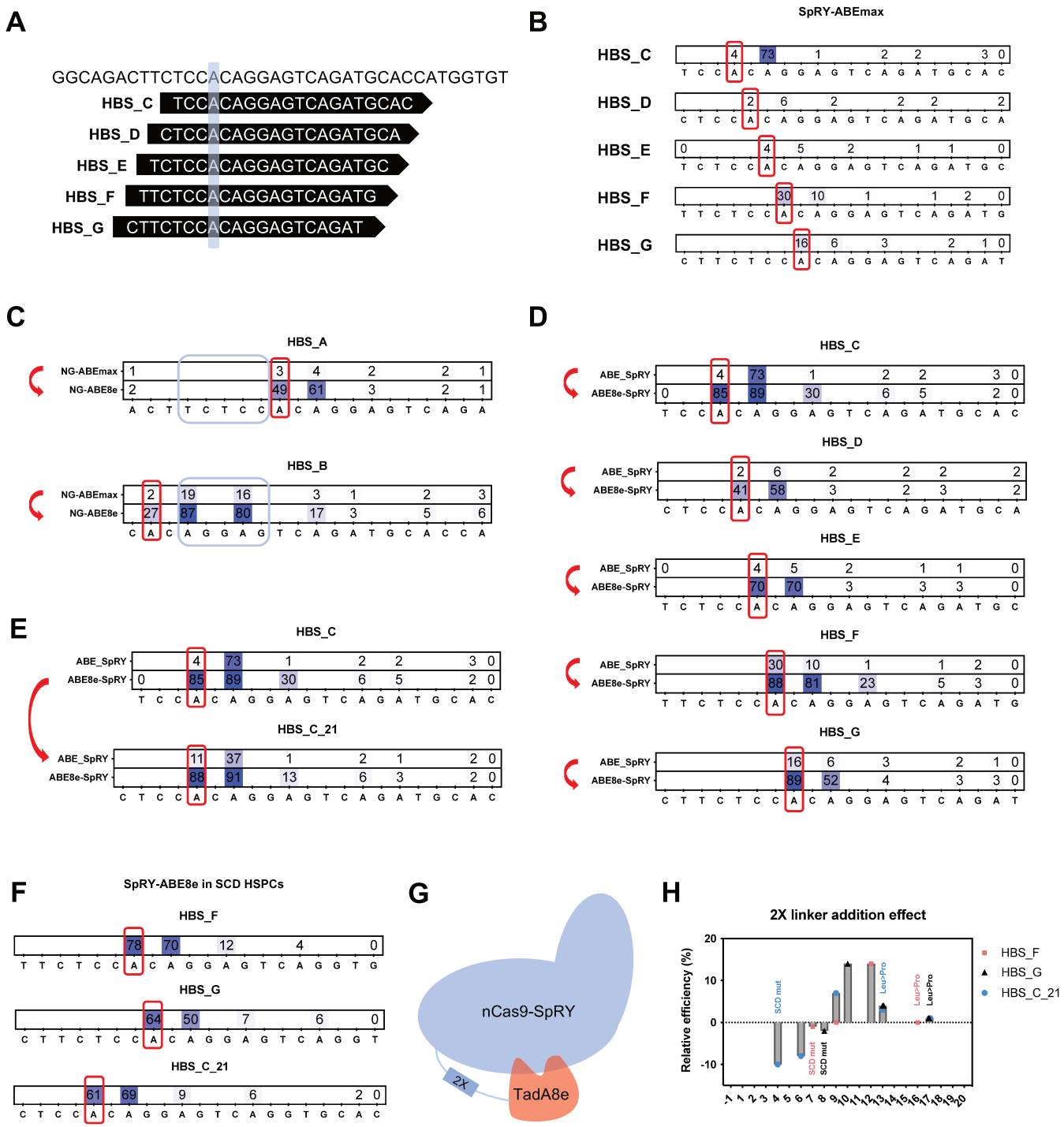

A
GGCAGACTTCTCCACAGGAGTCAGATGCACCATGGTGT
HBS_C TCCACAGGAGTCAGATGCAC
HBS_D CTCCACAGGAGTCAGATGCA
HBS_E TCTCCACAGGAGTCAGATGC
HBS_F TTCTCCACAGGAGTCAGATG
HBS_G CTTCTCCACAGGAGTCAGAT
B
SpRY-ABEmax
HBS_C
HBS_D
HBS_E
HBS_F
HBS_G
C
HBS_A
NG-ABEmax
NG-ABE8e
HBS_B
NG-ABEmax
NG-ABE8e
D
HBS_C
ABE_SpRY
ABE8e-SpRY
HBS_D
HBS_E
HBS_F
HBS_G
E
HBS_C
ABE_SpRY
ABE8e-SpRY
HBS_C_21
F
SpRY-ABE8e in SCD HSPCs
HBS_F
HBS_G
HBS_C_21
G
nCas9-SpRY
2X
TadA8e
H
2X linker addition effect
Relative efficiency (%)
SCD mut
Leu>Pro
HBS_F
HBS_G
HBS_C_21

Figure 3

E. Figure legends

Figure 1. HbG-Makassar cannot polymerize.

A. Structure of the amino acids at position 6 of the WT, sickle and Makassar β-globin, respectively. **B.** Bioinformatic prediction of the structure of the inter-β-strand interface between two deoxy-HbS (black and red panels) or two deoxy-HbG (blue panel).

Figure 2. Optimization of a base editing strategy to create the HbG-Makassar mutation in SCD-HUDEP2 using extended gRNAs and ABE linkers.

A. gRNAs HBS_A-B were manually designed according to available canonical PAMs available near to the SCD mutation. The predicted editing window is highlighted with a blue box while the target base position is indicated in red. NG and NGG PAMs are underlined. **B.** Base editing efficiencies obtained in SCD-HUDEP2 using NG-ABEmax, xABEmax or ABEmax ABEs and HBS_A or HBS_B gRNAs. The predicted editing window is highlighted with a blue box while the target base position is indicated in red (n=2 independent experiments). **C.** Extension of the gRNA could further unwind targeted DNA and allow the deaminase to access bases on the opposite side of the PAM. **D.** Base editing efficiencies obtained from positions 1 to 8 of the target DNA in SCD-HUDEP2 transfected with ABEmax, NG-ABEmax or xABEmax in parallel with HBS_B or its extended versions (n=2 independent experiments). **E.** Modification of the linker as a strategy to extend the editing window toward the PAM. Either the ssDBD Rad51 or 2 NLS sequences (2X) were included in the ABE linker sequence. **F.** Base editing efficiencies obtained from positions 1 to 20 of the target DNA in SCD-HUDEP2 transfected with NG-ABEmax-2X (left panel) or NG-ABEmax-Rad51 (right panel) and HBS_A (n=2 independent experiments). **G.** Base editing efficiencies obtained from positions 1 to 20 of the target DNA in SCD-HUDEP2 transfected with NG-ABEmax-2X and HBS_B (n=2 independent experiments).

Figure 2. Optimization of a base editing strategy to create the HbG-Makassar mutation in SCD-HUDEP2 and HSPCs using the PAM-less Cas9 SpRY, the highly processive deaminase ABE8e, an extended gRNA and an ABE linker.

A. gRNAs HBS_C to G were manually designed to place the SCD mutation within the predicted editing window (positions 4-8). The SCD mutation is highlighted with a blue box. **B.** Base editing efficiencies obtained from positions 1 to 20 of the target DNA in SCD-HUDEP2 transfected with SpRY-ABEmax and HBS_C to G (n=2 independent experiments). **C.** Base editing efficiencies obtained from positions 1 to 20 of the target DNA in SCD-HUDEP2 transfected with NG-ABE8e and HBS_A or HBS_B. (n=2 independent experiments). **D.** Base editing efficiencies obtained from positions 1 to 20 of the target DNA in SCD-HUDEP2 transfected with ABE8e-SpRY and HBS_C to G (n=2 independent experiments). **E.** Base editing efficiencies obtained from positions 1 to 20 of targeted DNA in SCD-HUDEP2 transfected with ABE8e-SpRY and HBS_C or HBS_C_21. (n=2 independent experiments). **F.** Base editing efficiencies obtained from positions 1 to 20 of the target DNA in SCD-HSPCs transfected with ABE8e-SpRY and HBS_F, G and C_21 (n=1 experiment).
G. Schematic representation of the ABE8e-SpRY-2X base editor. This enzyme contains the PAM-less Cas9 SpRY, an extended linker containing two additional NLS sequences (2X), and the highly processive deaminase TadA8e. **H.** Base editing efficiencies obtained using ABE8e-SpRY-2X enzyme in combination with HBS_F, G and C_21 gRNAs in SCD HSPCs relative to the values obtained under ABE8e-SpRY conditions. Relative efficiencies are expressed as percentages and calculated as following: relative efficiency = efficiency (ABE8e-SpRY-2X) - efficiency (ABE8e-SpRY). (n=1 experiment). SCD mut = SCD mutation position, Leu>Pro = bystander n+9 position leading to a missense mutation.

F. Materials and Methods

gRNA design and production

We manually designed gRNAs targeting the SCD mutation (**Table 1**). For gRNA expression in SCD-HUDEP2, oligonucleotide duplexes containing the gRNA protospacers were ligated into Bbs I–digested MA128 plasmid (provided by M. Amendola, Genethon, France). For SCD HSPCs, we used chemically modified synthetic single gRNAs harboring 2′-O-methyl analogs and 3′-phosphorothioate nonhydrolyzable linkages at the first three 5′ and 3′ nucleotides (Synthego).

Base editor-expressing plasmids

Plasmids used in this study include ABEmax (Addgene, 112101), NG-ABEmax (Addgene, 124163), xABEmax (Addgene, 119813), SpRY-ABEmax (Addgene, 140003), NG-ABE8e (Addgene, 138491) and ABE8e (Addgene, 138489). NG-ABEmax-2X plasmid was created by inserting of the -2X motif within the linker sequence of NG-ABEmax as previously described by Zafra et al.[84]. NG-ABEmax-Rad51 plasmid was created by inserting of the Rad51 ssDBD within the linker sequence of NG-ABEmax as previously described by Zhang et al.[83]. ABE8e-SpRY plasmid was created by replacing the Cas9 sequence of the ABE8e plasmid by the SpRY-P2A-EGFP from the SpRY-ABEmax plasmid. ABE8e-SpRY-2X plasmid was created by inserting of the -2X motif within the linker sequence of ABE8e-SpRY as previously described by Zafra et al.[84].

mRNA *in vitro* transcription

ABE8e-SpRY plasmid was digested overnight with SapI (Thermo). The linearized plasmid was purified (QIAGEN, 28106) and eluted in DNase/RNase-free water. The *in vitro* transcription protocol (Invitrogen, AM1334) was modified as follows. The GTP nucleotide solution was used at a final concentration of 3 mM and the anti-reverse cap analog N7-Methyl-3'-O-Methyl-Guanosine-5'-Triphosphate-5'-Guanosine (Trilink, N-7003) was added at a final concentration of 12 mM. The incubation time was reduced to 30 minutes and followed by polyadenylation (Invitrogen, AM1350). mRNA was precipitated using LiCl.

Cell line culture and transfection

HUDEP2 cells[227] were cultured as previously described by Canver et al.[228]. HUDEP2 cells were transfected with 3.6 µg of a base editor–expressing plasmid and 1.2 µg of a gRNA-containing plasmid, unless otherwise stated. For base editor plasmids that do not express GFP, we co-transfected 250 ng of a GFPmax-expressing plasmid (Lonza). We used AMAXA Cell Line Nucleofector Kit V (VCA-1003) and L-29 programs (Nucleofector II). GFP^+ HUDEP2 cells were sorted using SH800 Cell Sorter (Sony Biotechnology).

Cell purification, culture and transfection

We obtained human peripheral blood non-mobilized $CD34^+$ HSPCs from SCD patients. Written informed consent was obtained from all subjects. All experiments were performed in accordance with the Declaration of Helsinki. The study was approved by the regional investigational review board (reference: DC 2014-2272, CPP Ile-de-France II "Hôpital Necker-Enfants malades"). HSPCs were purified using the CD34 MicroBead Kit (Miltenyi Biotec). Forty-eight hours before transfection, $CD34^+$ cells were thawed and cultured at a concentration of $5x10^5$ cells/mL in "HSPC medium" containing StemSpan (STEMCELL Technologies) supplemented with penicillin/streptomycin (Gibco), StemRegenin1 (STEMCELL Technologies) and the following recombinant human cytokines (PeproTech): human stem cell factor, Fms-like tyrosine kinase receptor 3 ligand, thrombopoietin and interleukin-3.

For HSPCs transfection, $2x10^5$ HSPCs per condition were transfected with 3.0 µg of the ABE-encoding mRNA and 3.2 µg of the synthetic gRNA using the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the CA-137 program (Nucleofector 4D).

Evaluation of editing efficiency

Genomic DNA was extracted (Invitrogen, K182002 or Zymo Research, ZD7001) from HUDEP2 cells >3 days after sorting or at day 6 after HSPCs transfection. To evaluate base editing efficiency at gRNA target sites, we amplified the *HBB* gene with specific primers (**Table 2**), followed by Sanger sequencing and EditR analysis[180].

RP-HPLC analysis of globin chains

RP-HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). Globin chains were separated by HPLC using a 250 mm×4.6 mm, 3.6-µm Aeris Widepore column (Phenomenex). Samples were

eluted with a gradient mixture of solution A (water/acetonitrile/trifluoroacetic acid, 95:5:0.1) and solution B (water/acetonitrile/trifluoroacetic acid, 5:95:0.1). The absorbance was measured at 220 nm.

gRNA	Sequence (5' to 3')	PAM sequence (5' to 3')	Position (hg19)	Strand	Compatible enzymes
HBS_A	ACTTCTCCACAGGAGTCAGA	NG	chr11:5248224-5248243 (*HBB*)	+	NG-ABEmax, xABEmax
HBS_B	CACAGGAGTCAGATGCACCA	NGG	chr11:5248231-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_B_21	CCACAGGAGTCAGATGCACCA	NGG	chr11:5248230-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_B_22	TCCACAGGAGTCAGATGCACCA	NGG	chr11:5248229-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_B_23	CTCCACAGGAGTCAGATGCACCA	NGG	chr11:5248228-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_B_25	TTCTCCACAGGAGTCAGATGCACCA	NGG	chr11:5248226-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_B_27	ACTTCTCCACAGGAGTCAGATGCACCA	NGG	chr11:5248224-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_B_29	AGACTTCTCCACAGGAGTCAGATGCACCA	NGG	chr11:5248222-5248250 (*HBB*)	+	ABEmax, NG-ABEmax, xABEmax
HBS_C	TCCACAGGAGTCAGATGCAC	NRN	chr11:5248229-5248248 (*HBB*)	+	SpRY-ABEmax
HBS_C_21	CTCCACAGGAGTCAGATGCAC	NRN	chr11:5248228-5248248 (*HBB*)	+	SpRY-ABEmax
HBS_D	CTCCACAGGAGTCAGATGCA	NYN	chr11:5248228-5248247 (*HBB*)	+	SpRY-ABEmax
HBS_E	TCTCCACAGGAGTCAGATGC	NRN	chr11:5248227-5248246 (*HBB*)	+	SpRY-ABEmax
HBS_F	TTCTCCACAGGAGTCAGATG	NRN	chr11:5248226-5248245 (*HBB*)	+	SpRY-ABEmax

HBS_G	CTTCTCCACAGGAGTCAGAT	NYN	chr11:5248225-5248244 (*HBB*)	+	SpRY-ABEmax
HBS_F_SNP	TTCTCCACAGGAGTCAGGTG	NRN	chr11:5248226-5248245 (*HBB*)	+	SpRY-ABEmax
HBS_G_SNP	CTTCTCCACAGGAGTCAGGT	NYN	chr11:5248225-5248244 (*HBB*)	+	SpRY-ABEmax
HBS_C_21_SNP	CTCCACAGGAGTCAGGTGCAC	NRN	chr11:5248228-5248248 (*HBB*)	+	SpRY-ABEmax

Table 1. gRNA sequences.

Primer name	**Sequence (5'-3')**
HBB_F	CAGATCCCCAAAGGACTCAA
HBB_R	ACTCCTAAGCCAGTGCCAGA

Table 2. Primer sequences.

VI. Discussion and perspectives

β-hemoglobinopathies are the most common inherited monogenic disorders and among the most devastating diseases in the world. The lack of a definitive treatment available for all the patients, the accessibility of the target stem cell population to correct and the good comprehension of genetic mechanisms underlying this group of diseases justified the early development of gene therapy strategies for these diseases.

The first gene therapy strategies were based on gene addition through lentiviral transduction of HSCs and subsequent expression of a therapeutic Hb. These strategies have been very successful in the treatment of SCD and β-thalassemia[229]. However, they have some limitations in terms of efficacy and safety. On the one hand, their efficacy has been shown to be limited in case of a low vector copy number per cell associated with insufficient therapeutic globin expression[118]. On the other hand, the semi-random integration of lentiviral vectors is associated with a potential genotoxic risk[230].

A. Genome editing approaches for β-hemoglobinopathies

Recent advances in the genome editing field raised the possibility to develop gene therapies by inserting direct modifications in specific endogenous loci, thus permitting to reach higher levels of therapeutic Hb expression and avoid the potential genotoxicity resulting from lentiviral vector integration. These genome editing strategies can be classified in mutation-independent and mutation-dependent.

Most of mutation-independent approaches aim to reactivate HbF production. This strategy is interesting because it is validated by human genetics, mimicking a condition known to prevent the onset of symptoms in β-thalassemia and SCD patients[231]. Moreover, it has the advantage of being universal, and applicable regardless of the patient's genotype - which is of particular interest for β-thalassemia that can be cause by more than 350 mutations. Today, the most advanced clinical trial based on genome editing to reactivate HbF shows edifying results[112]. However, the benefit, and in particular the persistence of therapeutic levels of HbF expression

in the long term, remains to be evaluated. Moreover, concerns were raised regarding the impact of elevated HbF levels, especially in terms of growth retardation of fetuses from women with very high HbF levels[177]. Another category of mutation-independent strategies consists in α-globin downregulation. Several strategies aiming to disrupt the α-globin enhancer or to delete one of the *HBA* genes, demonstrated substantial correction of the α/non-α ratio in erythroid cells derived from β-thalassemia HSPCs[232–234].

Strategies aiming to correct specific mutations require each of them a specific optimization. However, in the case of β-thalassemia, they have the advantage of preserving the physiological regulation of genes, potentially restoring a high level of expression of HbA. Furthermore, in the case of SCD, these strategies can directly reduce the production of HbS, whose cellular concentration is the main factor responsible for polymerization[235].

Nevertheless, the vast majority of these approaches rely on the use of nucleases that can generate genotoxic DSBs. More recently, the same strategies are being explored using DSB-free genome editing tools such as base editors, which can be highly efficient in non-dividing HSCs[126,127]. It is in this context that we aimed to develop novel and safer base editing strategies.

Our team first focused on the development of safe base editing strategies aiming to induce HbF for the universal treatment of β-hemoglobinopathies. In particular, we used base editors to dissect the -200 region of *HBG* promoters and to identify critical base conversions that lead to high HbF levels. Our work identified KLF1 activator binding site creation as a potent strategy leading to therapeutically relevant HbF levels that rescued both the SCD and the β-thalassemia phenotypes, with editing of a large proportion of repopulating HSCs (**Annex 1**).

My main work was focused on the development of gene correction approaches using base editors.

Regarding SCD, base editing allows the conversion of the unique disease-causing mutation into a non-pathogenic allelic variant, giving rise to HbG-Makassar. While the

SCD mutation is difficult to access with first-generation base editors, we developed a strategy to efficiently reach this base, and obtained ~80% mutation correction in SCD HSPCs after applying different structural evolutions to ABEs. Moreover, our future work aims to further describe the structure and physico-chemical properties of this variant. Importantly, we have already obtained RBCs pellets from a HbG-Makassar carrier, that we will use to extract and purify this Hb variant. Then, we plan to submit it to different tests in order to describe its oxygen affinity, allosteric behavior, stability and determine its susceptibility to polymerization. This study will help to evaluate the clinical feasibility of such a gene correction strategy for the treatment of SCD patients.

Whereas SCD always results from a single nucleotide substitution, β-thalassemia can result from a large variety of mutations including substitutions, small insertions and deletions. Among the most common mutations found in the Mediterranean area and Middle East, we identified three severe mutations amenable to correction via base editing. For each of them, we developed a specific strategy and demonstrated its efficacy to correct the β-thalassemia phenotype *in vitro*. Of note, two base editing strategies were recently developed to correct two different β-thalassemia mutations, that are prevalent in Asia[126,236].

B. Editing in long-term repopulating HSCs

Importantly, clinical translation of these approaches will require additional efficacy and safety evaluation. In fact, while ABE-based strategies developed in this thesis showed a high efficiency in HSPCs, similarly to other reported studies[127,128](**Annex 1**), it is important to evaluate the extent and impact of base editing in repopulating HSCs. Hence, xenotransplantation experiments are required for evaluating the frequency of engrafting base-edited HSCs and their differentiation capability. Previous studies indicate that ABE-edited human HSPCs preserve multipotent capacities and efficiently repopulate the bone marrow of immunodeficient mice[127,128]. However, they showed the reduction of editing rates over time after transplant[127,128]. The observed reduction might result from lower editing efficiency in repopulating HSCs within the complex mixture of $CD34^+$ cells, or from a negative impact of editing on HSCs stemness or viability. In order to predict graft intake, maintenance and

differentiation capability of treated HSCs over time, SCD and β-thalassemia human $CD34^+$ cells edited with our strategies will be transplanted into immunodeficient mice. Such *in vivo* studies will be performed for the 4 mutation-correction strategies developed in this thesis, and completed by a secondary xenotransplantation experiment to confirm the long-term repopulating potential of corrected HSCs. Preliminary data using IVS1-110 HSPCs, showed 75% of editing efficiency 8 weeks post-transplantation.

C. Off-target editing

In addition, base editors can cause substantial off-target editing in both genomic DNA and RNA. Thus, it is essential to search for, quantify and characterize these events before clinical translation of a new gene therapy strategy. In our study, we nominated potential gRNA-dependent off-targets by *in silico* prediction and validated them in the target cells by targeted amplicon NGS. Correction of the IVS1-110 (G>A) mutation was associated with a low off-target activity, with no predicted impact on cell function. While these results are encouraging regarding the safety of our approaches, similar analyses need to be performed for the three other gene correction strategies. To complete this analysis, unbiased genome wide off-target nomination will be performed empirically by GUIDE-seq[222] and potential additional off-target sites will be validated by targeted amplicon NGS in edited cells. Moreover, gRNA-independent DNA and RNA off-targets will be screened empirically by whole exome sequencing (WES) and RNA-seq of base edited HSPCs. If these analyses reveal potentially harmful DNA or RNA off-targets, structural changes of the enzymes could be implemented in order to limit the frequency of these events. Indeed, replacement of the nCas9 by high-fidelity variants is an efficient way to reduce gRNA-dependent off-targets[93,94]. On the other hand, gRNA-independent DNA and RNA off-target activity can be reduced by the usage of deaminase variants carrying specific mutations[82]. Finally, as the overall off-target load significantly depends on exposure time of HSCs to the base editing system, RNP usage would be an efficient way to limit this risk. However, its usage is limited as base editor proteins production remains challenging[124]. However, it is important to notice that while these strategies are efficient ways to control the off-target activity, they could possibly interfere with the on-target activity.

D. Bystander editing

Bystander edits are an important challenge to face with base editing. Depending on the target sequence, we described bystander edits corresponding to intronic variants, silent mutations or missense mutations. The effect of intronic variants, as for instance in the case of IVS1-110 (G>A) mutation correction, was evaluated by assessing the impact of these edits in WT cells. No significant alterations were observed at mRNA or protein level. On the other hand, we observed silent and missense mutations in the *HBB* coding sequence while correcting β^0 and β^S mutations. Known coding mutations in the *HBB* sequence are reported and documented in the IthaGenes database, which allowed us to identify for each of them a previously described Hb variant carrying the relevant amino acid change and to anticipate its functionality[237]. Finally, the effect of silent mutations is less documented. Even though they are generally presumed to be neutral, recent studies highlighted the impact that they could have on protein expression[85,86]. In particular, replacement of a codon by a redundant one could alter mRNA structure, stability and translation efficacy[85]. Careful evaluation of these parameters should be performed to complete our studies on the correction of β^0 and β^S mutations.

E. Prime editing

While base editors are limited by the limited number of possible base conversions (A>G or C>T), a new technology called "prime editing" was recently developed[225]. Prime editing is based on a new enzyme, combining an inactivated Cas9 and a reverse transcriptase, used together with a prime editing gRNA (pegRNA) that includes a reverse transcription template containing the desired edits. After single-strand DNA cleavage at the targeted locus, this template is retrotranscribed by the enzyme, thus allowing the insertion of the desired modifications. Any modification (insertion, deletion or mutation of one or more base pairs) can be precisely inserted without requiring a DSB. Although this technology is currently showing limited efficacy in primary cells[226], it has allowed the precise correction of the SCD in a cell line[225] and could allow the correction of a wider variety of β-thalassemia mutations. Thus, further development of this strategy could offer a possible alternative to base

editing, with the added benefit of avoiding bystander mutations, which can restrict base editing application in some loci.

F. Expected phenotypic correction using these approaches

Importantly, strategies aiming to correct β-thalassemia mutations presented in this thesis were tested both in homozygous and compound heterozygous thalassemic HSPCs. It appeared that correction of the β^+ IVS1-110 (G>A) mutation on only one of the alleles was slightly less efficient in terms of amelioration of the α/non-α ratio as compared to the correction of both alleles. This finding was not observed when comparing the phenotypic improvement obtained after biallelic or monoallelic correction of the β^0 CD39 (CAG>TAG) mutation. This difference is probably due to the variability of the genetic background of the donors used in this study. Future works will also aim to evaluate the correction of the SCD phenotype in the context of sickle β-thalassemia. In fact, correction of the β^+ or β^0 mutation inherited in parallel with the β^S one would be the equivalent of generating the genotype of an asymptomatic individual heterozygous for the SCD mutation. Finally, one of the key advantages of base editing also resides in the fact that it can combine two targets. In fact, as it does not generate DSBs, concomitant editing of two genomic sites theoretically does not lead to genomic rearrangements (such as translocations) and should not increase toxicity as compared to nuclease-based genome editing. Thus, we envision that base editing strategies developed in this thesis could be used in parallel, or in combination with strategies aiming to increase HbF production or downregulate α-globin. One study already evaluated the feasibility to combine gene correction and HbF reactivation in β-thalassemic HSPCs, showing an additive effect in the correction of the phenotype[126].

Experience from mixed chimerism following allogenic bone marrow transplant shows that healthy erythroblasts undergo *in vivo* selection, allowing the resolution of β-thalassemia with a proportion of healthy donor cells as low as 20-30%[238,239] and the correction of the SCD phenotype with 30% of donor chimerism[240]. Therefore, correction efficiencies obtained with our strategies substantially exceeded these thresholds, maybe even considering (i) the potential decrease in correction efficiency

after *in vivo* engraftment and (ii) a potential reduction in on-target efficiency if base editor modifications are required following the RNA/DNA off-targets screening.

Strategies presented in this thesis add up to the numerous approaches in development or already in clinics, each having its own advantages and disadvantages. In the area of precision medicine, it is interesting to note that unequally effective results have been reported among SCD or β-thalassemia patients after long-term follow up of patients included in clinical trials. The possibility was raised that the genetic mutations responsible for the disease (β^s/β^s or β^0/β^s for SCD[213], or the combination of β^+/β^0 mutations for β-thalassemia[118]) or even the genetic background of the patient could affect the clinical outcome of the treatment[213]. In fact, the long-term experience in SCD or β-thalassemia clinical management has put in evidence the impact of some disease-modifying genes on treatment response[50,241]. Thus, on the basis of future clinical trial results, it is likely that these different gene therapy treatments will be complementary one to each other, and indicated to different patients' subsets stratified by their genotype and polymorphisms.

G. The challenge of costs and accessibility to gene therapy

Moving these strategies into routine care is a challenge on several levels. On the one hand, cost is a significant limitation for these gene therapy strategies. Recently, the marketing of Zynteglo®, the only gene therapy approved in Europe for the treatment of β-thalassemia (based on the use of a lentiviral vector), was suspended following the failure to complete negotiations with European payers for the coverage of the treatment proposed for $1.8 million by the company[242,243]. In the US, this treatment was recently approved and will come at a price of $2.8 million. Interestingly, the company engaged to reimburse up to 80% of the cost of the therapy if a patient fails to achieve and maintain transfusion independence up to two years following infusion[244]. Such facts highlight the need to develop new coverage systems for healthcare to facilitate patients access to innovative treatments.

The costs of these definitive therapies should be compared with the annual cost of regular caring for patients, and the gain in quality of life for treated patients. In fact, a

recent study estimated that people with SCD spend roughly $1.7 million on total lifetime medical costs[245] while Bluebird Bio claimed that for a patient with transfusion-dependent β-thalassemia it can reach up to $6.4 million in the U.S. However, to date, the lack of long-term studies makes the evaluation of gene therapy benefits uncertain. The rigorous and shared collection of efficacy and safety data will facilitate the calculation of the cost-effectiveness ratio[246]. Finally, as it is estimated that more than 75% of newborns with SCD are born in sub-Saharan Africa, solutions that would allow as many people as possible to benefit from these therapies must be studied. Indeed, the units allowing HSCs transplantation are either unavailable or not sufficient in most of the regions in which SCD is very prevalent[247].

The new genome editing approaches based on the use of CRISPR/Cas9 and derived tools are free of the cost of lentivirus manufacturing, which represents approximately 50% of the cost of the production for therapies based on these vectors[248]. Thus, their cost should decrease even if to date the production of reagents (gRNA, Cas9 and base editors in the form of mRNA or protein) of clinical grade and on a large scale remains a laborious and costly process[249]. The development of an increasing number of gene therapies in various fields should make it possible to reduce the cost of production of these reagents and thus make these new treatments more accessible. Alternatively, the reduction in the cost of reagents could be achieved by reducing the number of transfected cells. To achieve this, different approaches are being evaluated, such as the isolation and transfection of a very pure population of HSCs (rather than a mixture of HSCs and progenitors)[153,250].

Furthermore, the process of HSPC collection and toxic myeloablation remains an invasive treatment for the patient and represents a barrier to the widespread application of these approaches. The procedure requires several days of drug mobilization allowing the isolation of HSPCs by leukapheresis followed by myeloablative conditioning and transplantation of the treated cells. The myeloablative treatment, in addition to prolonging hospitalization, is associated with loss of fertility and may result in secondary cancers. Indeed, it has been suggested that this treatment has been responsible for the myelodysplasia observed after gene therapy in a SCD patient[251]. The development of an *in vivo* gene therapy strategy would simplify this clinical protocol. In addition, it would eliminate the demanding process of

ex vivo culture steps. For current therapies, these steps are a limit to efficacy because they lead to a significant loss in HSCs viability and engraftment capacity[252]. In addition, it is limiting for many centers that do not have the necessary facilities for cell culture. Interestingly, one team has shown the feasibility of intravenous injection of a non-integrating virus to edit previously mobilized HSCs in mice[253]. However, to date, this type of approach faces major challenges such as the immunogenicity of the vector, its biodistribution and its non-tissue specificity[254]. Finally,
another field of research aims to develop *in utero* stem cell transplantation in order to prevent the early clinical manifestations of SCD and β-thalassemia[255,256]. This cellular therapy takes advantage of the relative immune immaturity of the fetus during early stages of development to develop a long-term donor-specific tolerance. However, efficiency of this strategy is for now limited by the low levels of chimerism obtained[255]. Thus, another option is to directly proceed with *in utero* gene therapy to the fetus by injecting viral vectors or nanoparticles. A proof-of-principle study demonstrated that *in utero* gene therapy using lentiviral vectors corrected Hb levels and the pathological phenotype of a heterozygous humanized mouse model of β-thalassemia[257].

H. Conclusions

To conclude, this work demonstrates the efficacy of base editing approaches to correct prevalent and severe β-thalassemia mutations or to convert the SCD mutation to a known non-pathogenic Hb variant mutation. Further characterization of the off-target activity, and validation of these results in HSPCs *in vivo* will provide sufficient proof of efficacy and safety to enable the clinical development of base-edited HSPCs for gene therapy of β-hemoglobinopathies.

♥ A HUGE THANK YOU … UN GRAND MERCI … GRAZIE MILLE … ΕΥΧΑΡΙΣΤΩ ΠΟΛΥ … ♥

A Marina Cavazzana, merci infiniment d'avoir rendu tout cela possible. Merci pour votre bienveillance, votre accessibilité. Merci d'avoir cru en moi, de m'avoir fait confiance. Merci d'avoir porté un regard enthousiaste et avisé sur mon travail. Vous êtes une source d'inspiration, j'admire votre engagement auprès des patients, de la recherche, mais aussi l'attention particulière que vous portez à la réussite des personnes qui vous entoure. Merci!

To Annarita Miccio, I feel so lucky that I had the opportunity to work with you. Thank you for your guidance along this path. Your door was always open, and you would always help me to structure my mind, make me feel motivated and enthusiastic. Your strong scientific curiosity and critical mind inspired me so much. I admire how clever and patient you are, but also how tireless you are in achieving and progressing! Thank you so much.

To Mario Amendola, thank you for your continuous support along my journey. Thank you for contributing to all my thesis committee and always giving me advice. And finally, on this occasion, thank you for accepting to be part of my jury.

A Wassim El Nemer, pour m'avoir également accompagnée dans mon parcours. Merci pour ta présence lors de mes comités, pour ton soutien bienveillant et ton regard critique sur mes résultats.

To Paula Rio, John Strouboulis and Carine Giovannangeli for accepting to be part of my jury and taking on your time to evaluate my work.

A Nabila Bouatia-Naji, tu as été la première à me faire confiance, en M1, pour m'accueillir en stage. Tu m'as transmis ta passion enthousiaste pour la recherche et donné l'envie et la confiance pour poursuivre vers un doctorat. Merci infiniment.

A Stuart Edelstein, pour m'avoir accompagnée dans mes projets depuis plus de 10 ans maintenant. Si j'en suis ici aujourd'hui c'est aussi grâce à toi ; et ce travail aussi, car il s'appuie en partie sur tes résultats.

To Panagiotis… without you my PhD journey would not have been the same, for sure! I am so grateful for these years by your side, sharing crazy laughs and deep scientific discussions at the same time. Thank you for being so kind, patient, and always making me feel confident when I was not. I've learned so much from you, you've been a mentor to me and to other many students of the lab. I admire your strength, capacity of work while being so available and kind. I will miss you so much !

A Anne, parce que ta gentillesse et ta bienveillance ont été d'un réconfort inégalable pendant les moments difficiles mais aussi pour ta joie de vivre et ton humour qui ont égayés ces années ! Mais tu es aussi une scientifique courageuse et déterminée, une médiatrice dévouée et passionnée … Tellement de cordes à ton arc ! Courage pour ta fin de thèse, tu vas gérer ! Ps : Peut-être qu'un jour on arrivera enfin à manger un légume qu'on a planté … ? =D

Giacomo, it has been a pleasure working with you and sharing many good times with you. You're definitely a cool guy but also a kind and benevolent person that I've missed so much during my last year of PhD. I wish you all the best in your following experiences, anyone working with you is lucky to have you!

A Pierre, sans toi je n'y serais pas parvenue ! Merci beaucoup pour ton aide précieuse pour les manips, ton regard critique et tes conseils avisés. Mais aussi, je suis honorée d'avoir faire partie de tes premiers élèves de CrossFit, j'ai hâte de découvrir ta salle (ou pas =D) !

A Sophie, pour nous avoir montré le chemin ; merci pour tes conseils et tes remarques constructives concernant les manips, les présentations... Mais aussi pour tes petites attentions envers les autres, les fous rires dans le bureau et nos longues discussions :)

A Letizia, che bella esperienza lavorare con te. Sei intelligente, sempre volontaria, e piena di energia (anche dopo 10 ChIP =D). Grazie per avermi aiutata durante il mio secondo anno di tesi ma anche per la tua presenza gioviale di ogni giorno. Ti auguro tante belle cose, perché te lo meriti; e in bocca lupo per la tesi!

A Tristan, tout d'abord pour m'avoir accueillie et transmis tes connaissances durant mon stage de M2 à l'époque, mais aussi plus récemment pour ta contribution essentielle pour les manips *in vivo* ! Mais également pour ton humour atypique de tous les jours =D

A Mégane, pour tes nombreux conseils de post-doc avisée et ta disponibilité. Mais aussi pour avoir toujours eu plaisir à m'expliquer les subtilités de certains concepts nouveaux pour moi... Merci !

A Mathieu, aka MvJ. Merci pour ton support, tes petites astuces et pour être toujours à l'écoute ; à peine arrivé tu me coachais déjà sur le Fortessa ! J'ai hâte de suivre l'évolution de MvJ Therapeutics !

To the interns, and former engineers, for bringing joy and a fresh view on our work: Bochra, Jeanne, Zoe, Gaia, Priyanka, Sarah, Gilles, Marta…

To the Imagine Sport Association, and in particular to our presidential team - Tobias and Alice, but also to all the coaches, you did an amazing job to maintain us in shape!

A Elisa Magrin, pour m'avoir accompagnée dans l'achèvement de mes études de pharmacie. Merci pour ton suivi, tes conseils et nos discussions qui m'aident à mieux appréhender les aspects cliniques de notre travail, le tout dans la bonne humeur !

A Sandra Manceau et Laure Joseph, pour votre disponibilité, vos conseils et votre aide précieuse dans la mise en place de ce projet.

A mes amis d'enfance, de collège, lycée, fac ... pour tous les bons moments avec vous qui rendent la vie si belle et qui m'ont sans aucun doute aidée à traverser dette étape !

A mes parents, grands-parents, ma famille et Manu pour leur amour et soutien inconditionnel.

Et finalement, aux patients qui ont acceptés de prendre part à ces travaux, en espérant que la recherche scientifique puisse continuer à améliorer la prise en charge des beta-hémoglobinopathies.

Bibliography

1. Mettananda S. Ameliorating β-thalassaemia by manipulating expression of the α-globin gene. 2015;
2. Cavazzana M, Antoniani C, Miccio A. Gene Therapy for β-Hemoglobinopathies. *Mol. Ther.* 2017;25(5):1142–1154.
3. Voon HPJ, Vadolas J. Controlling α-globin: a review of α-globin expression and its impact on β-thalassemia. *Haematologica*. 2008;93(12):1868–1876.
4. Song S-H, Hou C, Dean A. A Positive Role for NLI/Ldb1 in Long-Range β-Globin Locus Control Region Function. *Mol. Cell*. 2007;28(5):810–822.
5. Thein SL, Menzel S. Discovering the genetics underlying foetal haemoglobin production in adults. *Br. J. Haematol.* 2009;145(4):455–467.
6. Powars DR, Weiss JN, Chan LS, Schroeder WA. Is there a threshold level of fetal hemoglobin that ameliorates morbidity in sickle cell anemia? *Blood*. 1984;63(4):921–926.
7. Forget BG. Molecular Basis of Hereditary Persistence of Fetal Hemoglobin. *Ann. N. Y. Acad. Sci.* 1998;850(1):38–44.
8. Sankaran VG, Xu J, Byron R, et al. A functional element necessary for fetal hemoglobin silencing. *N. Engl. J. Med.* 2011;365(9):807–814.
9. Martyn GE, Wienert B, Yang L, et al. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat. Genet.* 2018;50(4):498–503.
10. Masuda T, Wang X, Maeda M, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science*. 2016;351(6270):285–289.
11. Wienert B, Martyn GE, Kurita R, et al. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood*. 2017;130(6):803–807.
12. Wienert B, Funnell APW, Norton LJ, et al. Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat. Commun.* 2015;6:7085.
13. Martyn GE, Wienert B, Kurita R, et al. A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. *Blood*. 2019;133(8):852–856.
14. Martyn GE, Quinlan KGR, Crossley M. The regulation of human globin promoters by CCAAT box elements and the recruitment of NF-Y. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 2017;1860(5):525–536.
15. Loughran SJ, Haas S, Wilkinson AC, Klein AM, Brand M. Lineage commitment of hematopoietic stem cells and progenitors: insights from recent single cell and lineage tracing technologies. *Exp. Hematol.* 2020;88:1–6.
16. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* 2017;19(4):271–281.
17. Rodrigues CP, Shvedunova M, Akhtar A. Epigenetic Regulators as the Gatekeepers of Hematopoiesis. *Trends Genet.* 2021;37(2):125–142.
18. Zivot A, Lipton JM, Narla A, Blanc L. Erythropoiesis: insights into pathophysiology and treatments in 2017. *Mol. Med. Camb. Mass*. 2018;24(1):11.
19. Orkin SH. Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.* 2000;1(1):57–64.
20. Peter Klinken S. Red blood cells. *Int. J. Biochem. Cell Biol.* 2002;34(12):1513–1518.
21. Manwani D, Bieker JJ. The Erythroblastic Island. *Curr. Top. Dev. Biol.* 2008;82:23–53.
22. Chasis JA, Mohandas N. Erythroblastic islands: niches for erythropoiesis. *Blood*. 2008;112(3):470–478.

23. Cantú I, Philipsen S. Flicking the switch: Adult hemoglobin expression in erythroid cells derived from cord blood and human induced pluripotent stem cells. *Haematologica*. 2014;99:1647–9.
24. Piel FB, Patil AP, Howes RE, et al. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet Lond. Engl.* 2013;381(9861):142–151.
25. Weatherall D, Hofman K, Rodgers G, Ruffin J, Hrynkow S. A case for developing North-South partnerships for research in sickle cell disease. *Blood*. 2005;105(3):921–923.
26. Piel FB, Patil AP, Howes RE, et al. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat. Commun.* 2010;1(1):104.
27. Leleu H, Arlet JB, Habibi A, et al. Epidemiology and disease burden of sickle cell disease in France: A descriptive study based on a French nationwide claim database. *PloS One*. 2021;16(7):e0253986.
28. Herrick JB. Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. 1910. *Yale J. Biol. Med.* 2001;74(3):179–184.
29. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *The Lancet*. 2010;376(9757):2018–2031.
30. Reiter CD, Wang X, Tanus-Santos JE, et al. Cell-free hemoglobin limits nitric oxide bioavailability in sickle-cell disease. *Nat. Med.* 2002;8(12):1383–1389.
31. Orpha.net : La drépanocytose. Available at: https://www.orpha.net/data/patho/Pub/fr/Drepanocytose-FRfrPub125v01.pdf.
32. Piel FB, Steinberg MH, Rees DC. Sickle Cell Disease. *http://dx.doi.org.proxy.insermbiblio.inist.fr/10.1056/NEJMra1510865*. 2017;
33. HAS: Syndromes drépanocytaires majeurs de l'enfant et de l'adolescent. Protocole national de diagnostic et de soins pour une maladie rare. Janvier 2010.
34. McGann PT, Ware RE. Hydroxyurea therapy for sickle cell anemia. *Expert Opin. Drug Saf.* 2015;14(11):1749–1758.
35. Brewin J, El Hoss S, Strouboulis J, Rees D. A novel index to evaluate ineffective erythropoiesis in hematological diseases offers insights into sickle cell disease. *Haematologica*. 2021;107(1):338–341.
36. Ware RE. How I use hydroxyurea to treat young patients with sickle cell anemia. *Blood*. 2010;115(26):5300–5311.
37. Gladwin MT, Shelhamer JH, Ognibene FP, et al. Nitric oxide donor properties of hydroxyurea in patients with sickle cell disease. *Br. J. Haematol.* 2002;116(2):436–444.
38. Niihara Y, Matsui NM, Shen YM, et al. L-glutamine therapy reduces endothelial adhesion of sickle red blood cells to human umbilical vein endothelial cells. *BMC Blood Disord.* 2005;5:4.
39. Vichinsky E, Hoppe CC, Ataga KI, et al. A Phase 3 Randomized Trial of Voxelotor in Sickle Cell Disease. *N. Engl. J. Med.* 2019;381(6):509–519.
40. Bougouma EC, Sirima SB. Inherited Disorders of Hemoglobin and Plasmodium falciparum Malaria. IntechOpen; 2020.
41. Taher AT, Musallam KM, Cappellini MD. β-Thalassemias. *N. Engl. J. Med.* 2021;384(8):727–743.
42. Galanello R, Origa R. Beta-thalassemia. *Orphanet J. Rare Dis.* 2010;5(1):1–15.
43. Catherine Badens. La beta-Thalassémie en France : Données du Registre National.
44. Lechauve C, Keith J, Khandros E, et al. The autophagy-activating kinase ULK1 mediates clearance of free α-globin in β-thalassemia. *Sci. Transl. Med.* 2019;11(506):eaav4881.
45. Premawardhena A, Fisher CA, Olivieri NF, et al. A novel molecular basis for beta thalassemia intermedia poses new questions about its pathophysiology. *Blood*.

2005;106(9):3251–3255.
46. Khandros E, Thom CS, D'Souza J, Weiss MJ. Integrated protein quality-control pathways regulate free α-globin in murine β-thalassemia. *Blood*. 2012;119(22):5265–5275.
47. Wilairat P, Kittikalayawong A, Chaicharoen S. The thalassemic red cell membrane. *Southeast Asian J. Trop. Med. Public Health*. 1992;23 Suppl 2:74–78.
48. Vasilopoulou M, Stafylidis C, Politou M. The thrombotic spectrum of B-thalassemia. *Thromb. Update*. 2022;7:100102.
49. Origa R. β-Thalassemia. *Genet. Med.* 2017;19(6):609–619.
50. Cao A, Galanello R. Beta-thalassemia. *Genet. Med.* 2010;12(2):61–76.
51. HAS: Syndromes thalassémiques majeurs et intermédiaires. Protocole national de diagnostic et de soins pour une maladie rare. Juin 2008.
52. Cappellini MD, Taher AT. The use of luspatercept for thalassemia in adults. *Blood Adv.* 2021;5(1):326–333.
53. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346(6213):1258096.
54. Jinek M, Chylinski K, Fonfara I, et al. A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816–821.
55. Le Rhun A, Escalera-Maurer A, Bratovič M, Charpentier E. CRISPR-Cas in Streptococcus pyogenes. *RNA Biol.* 2019;16(4):380–389.
56. Mir A, Edraki A, Lee J, Sontheimer EJ. Type II-C CRISPR-Cas9 Biology, Mechanism and Application. *ACS Chem. Biol.* 2018;13(2):357–365.
57. Nishimasu H, Ran FA, Hsu PD, et al. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*. 2014;156(5):935–949.
58. Arora L, Narula A. Gene Editing and Crop Improvement Using CRISPR-Cas9 System. *Front. Plant Sci.* 2017;8:1932.
59. Genovese P, Schiroli G, Escobar G, et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*. 2014;510(7504):235–240.
60. Jang H-K, Song B, Hwang G-H, Bae S. Current trends in gene recovery mediated by the CRISPR-Cas system. *Exp. Mol. Med.* 2020;52(7):1016–1027.
61. Boutin J, Cappellen D, Rosier J, et al. ON-Target Adverse Events of CRISPR-Cas9 Nuclease: More Chaotic than Expected. *CRISPR J.* 2022;5(1):19–30.
62. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 2018;36(8):765–771.
63. Leibowitz ML, Papathanasiou S, Doerfler PA, et al. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* 2021;53(6):895–905.
64. Turchiano G, Andrieux G, Klermund J, et al. Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell*. 2021;28(6):1136-1147.e5.
65. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013;31(9):822–826.
66. Park SH, Lee CM, Dever DP, et al. Highly efficient editing of the β-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. *Nucleic Acids Res.* 2019;47(15):7955–7972.
67. Cromer MK, Vaidyanathan S, Ryan DE, et al. Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther. J. Am. Soc. Gene Ther.* 2018;26(10):2431–2442.
68. Nambiar TS, Baudrier L, Billon P, Ciccia A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell*. 2022;82(2):348–388.
69. Schiroli G, Conti A, Ferrari S, et al. Precise Gene Editing Preserves Hematopoietic

Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell*. 2019;24(4):551-565.e8.
70. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 2018;24(7):927–930.
71. Enache OM, Rendo V, Abdusamad M, et al. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* 2020;52(7):662–668.
72. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533(7603):420–424.
73. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 2020;38(7):824–844.
74. Gaudelli NM, Komor AC, Rees HA, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*. 2017;551(7681):464–471.
75. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 2018;19(12):770–788.
76. Gu S, Bodai Z, Cowan QT, Komor AC. Base Editors: Expanding the Types of DNA Damage Products Harnessed for Genome Editing. *Gene Genome Ed.* 2021;1:100005.
77. Huang TP, Zhao KT, Miller SM, et al. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* 2019;37(6):626–631.
78. Hu J, Miller S, Geurts M, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. 2018;556:.
79. Yang L, Zhang X, Wang L, et al. Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell*. 2018;9(9):814–819.
80. Miller SM, Wang T, Randolph PB, et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* 2020;38(4):471–481.
81. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*. 2020;368(6488):290–296.
82. Richter MF, Zhao KT, Eton E, et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 2020;38(7):883–891.
83. Zhang X, Chen L, Zhu B, et al. Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. *Nat. Cell Biol.* 2020;22(6):740–750.
84. Zafra MP, Schatoff EM, Katti A, et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* 2018;36(9):888–893.
85. Kabra M, Moosajee M, Newby GA, et al. Comprehensive Analysis of CRISPR Base Editing Outcomes for Multimeric Protein. 2022;2022.06.20.496792.
86. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature*. 2022;606(7915):725–731.
87. Kim HS, Jeong YK, Hur JK, Kim J-S, Bae S. Adenine base editors catalyze cytosine conversions in human cells. *Nat. Biotechnol.* 2019;37(10):1145–1148.
88. Jeong YK, Lee S, Hwang G-H, et al. Adenine base editor engineering reduces editing of bystander cytosines. *Nat. Biotechnol.* 2021;39(11):1426–1433.
89. Antoniou P, Miccio A, Brusson M. Base and Prime Editing Technologies for Blood Disorders. *Front. Genome Ed.* 2021;3:618406.
90. Porto EM, Komor AC, Slaymaker IM, Yeo GW. Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.* 2020;19(12):839–859.
91. Tu T, Song Z, Liu X, et al. A precise and efficient adenine base editor. *Mol. Ther.*

2022;30(9):2933–2941.
92. Vrtis KB, Dewar JM, Chistol G, et al. Single-strand DNA breaks cause replisome disassembly. *Mol. Cell*. 2021;81(6):1309-1318.e6.
93. Rees HA, Komor AC, Yeh W-H, et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* 2017;8(1):15790.
94. Lee JK, Jeong E, Lee J, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* 2018;9(1):3048.
95. Yu Y, Leete TC, Born DA, et al. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* 2020;11(1):2052.
96. Zuo E, Sun Y, Wei W, et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*. 2019;364(6437):289–292.
97. Doman JL, Raguram A, Newby GA, Liu DR. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* 2020;38(5):620–628.
98. Lee HK, Smith HE, Liu C, Willi M, Hennighausen L. Cytosine base editor 4 but not adenine base editor generates off-target mutations in mouse embryos. *Commun. Biol.* 2020;3(1):19.
99. Jin S, Zong Y, Gao Q, et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*. 2019;364(6437):292–295.
100. Grünewald J, Zhou R, Garcia SP, et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*. 2019;569(7756):433–437.
101. Rees HA, Wilson C, Doman JL, Liu DR. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* 2019;5(5):eaax5717.
102. Zhou C, Sun Y, Yan R, et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature*. 2019;571(7764):275–278.
103. Grünewald J, Zhou R, Iyer S, et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* 2019;37(9):1041–1048.
104. Zuo E, Sun Y, Yuan T, et al. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods*. 2020;17(6):600–604.
105. Gluckman E, Cappelli B, Bernaudin F, et al. Sickle cell disease: an international survey of results of HLA-identical sibling hematopoietic stem cell transplantation. *Blood*. 2017;129(11):1548–1556.
106. Eapen M, Brazauskas R, Walters MC, et al. Impact of donor type and conditioning regimen intensity on allogeneic transplantation outcomes in patients with Sickle Cell Disease: a retrospective, cohort study. *Lancet Haematol.* 2019;6(11):e585–e596.
107. Cavazzana M, Ribeil J-A, Lagresle-Peyrou C, André-Schmutz I. Gene Therapy with Hematopoietic Stem Cells: The Diseased Bone Marrow's Point of View. *Stem Cells Dev.* 2017;26(2):71–76.
108. Kim HK, De La Luz Sierra M, Williams CK, Gulino AV, Tosato G. G-CSF down-regulation of CXCR4 expression identified as a mechanism for mobilization of myeloid cells. *Blood*. 2006;108(3):812–820.
109. Greenbaum AM, Link DC. Mechanisms of G-CSF-mediated hematopoietic stem and progenitor mobilization. *Leukemia*. 2011;25(2):211–217.
110. Yannaki E, Karponi G, Zervou F, et al. Hematopoietic Stem Cell Mobilization for Gene Therapy: Superior Mobilization by the Combination of Granulocyte–Colony Stimulating Factor Plus Plerixafor in Patients with β-Thalassemia Major. *Hum. Gene Ther.* 2013;24(10):852–860.
111. Fitzhugh CD, Hsieh MM, Bolan CD, Saenz C, Tisdale JF. Granulocyte colony-

stimulating factor (G-CSF) administration in individuals with sickle cell disease: time for a moratorium? *Cytotherapy*. 2009;11(4):464–471.
112. Frangoul H, Altshuler D, Cappellini MD, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* 2021;384(3):252–260.
113. Lagresle-Peyrou C, Lefrère F, Magrin E, et al. Plerixafor enables safe, rapid, efficient mobilization of hematopoietic stem cells in sickle cell disease patients after exchange transfusion. *Haematologica*. 2018;103(5):778–786.
114. CD34 - Hematopoietic progenitor cell antigen CD34 - Homo sapiens (Human) | UniProtKB | UniProt.
115. Cavazzana-Calvo M, Payen E, Negre O, et al. Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature*. 2010;467(7313):318–322.
116. Negre O, Bartholomae C, Beuzard Y, et al. Preclinical evaluation of efficacy and safety of an improved lentiviral vector for the treatment of β-thalassemia and sickle cell disease. *Curr. Gene Ther.* 2015;15(1):64–81.
117. Ribeil J-A, Hacein-Bey-Abina S, Payen E, et al. Gene Therapy in a Patient with Sickle Cell Disease. *http://dx.doi.org.proxy.insermbiblio.inist.fr/10.1056/NEJMoa1609677*. 2017;
118. Thompson AA, Walters MC, Kwiatkowski J, et al. Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. *N. Engl. J. Med.* 2018;
119. Kwiatkowski JL, Walters MC, Hongeng S, et al. Long-Term Efficacy and Safety of Betibeglogene Autotemcel Gene Therapy for the Treatment of Transfusion-Dependent β-Thalassemia: Results in Patients with up to 6 Years of Follow-up. *Blood*. 2020;136(Supplement 1):51–52.
120. Malik P, Grimley M, Quinn CT, et al. Gene Therapy for Sickle Cell Anemia Using a Modified Gamma Globin Lentivirus Vector and Reduced Intensity Conditioning Transplant Shows Promising Correction of the Disease Phenotype. *Blood*. 2018;132(Supplement 1):1021–1021.
121. Grimley M. Early Results from a Phase 1/2 Study of Aru-1801 Gene Therapy for Sickle Cell Disease (SCD): Manufacturing Process Enhancements Improve Efficacy of a Modified Gamma Globin Lentivirus Vector and Reduced Intensity Conditioning Transplant. 2020;
122. Marktel S, Scaramuzza S, Cicalese MP, et al. Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent ß-thalassemia. *Nat. Med.* 2019;25(2):234–241.
123. Esrick EB, Lehmann LE, Biffi A, et al. Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease. *N. Engl. J. Med.* 2021;384(3):205–215.
124. Jang H-K, Jo DH, Lee S-N, et al. High-purity production and precise editing of DNA base editing ribonucleoproteins. *Sci. Adv.* 2021;7(35):eabg2661.
125. Lattanzi A, Meneghini V, Pavani G, et al. Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements. *Mol. Ther.* 2019;27(1):137–150.
126. Zeng J, Wu Y, Ren C, et al. Therapeutic base editing of human hematopoietic stem cells. *Nat. Med.* 2020;26(4):535–541.
127. Newby GA, Yen JS, Woodard KJ, et al. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature*. 2021;595(7866):295–302.
128. Siegner SM, Clemens A, Ugalde L, et al. Adenine base editing is an efficient approach to restore function in FA patient cells without double-stranded DNA breaks. 2022;2022.04.22.489197.
129. Ravi NS, Wienert B, Wyman SK, et al. Identification of novel HPFH-like mutations by CRISPR base editing that elevate the expression of fetal hemoglobin. *eLife*. 2022;11:e65421.

130. Dever DP, Bak RO, Reinisch A, et al. CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature*. 2016;539(7629):384–389.
131. DeWitt MA, Magis W, Bray NL, et al. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* 2016;8(360):360ra134-360ra134.
132. Lattanzi A, Camarena J, Lahiri P, et al. Development of β-globin gene correction in human hematopoietic stem cells as a potential durable treatment for sickle cell disease. *Sci. Transl. Med.* 2021;13(598):eabf2444.
133. UC Consortium Launches First Clinical Trial Using CRISPR to Correct Gene Defect That Causes Sickle Cell Disease. *UC Consort. Launches First Clin. Trial Using CRISPR Correct Gene Defect Causes Sick. Cell Dis. UC San Franc.* .
134. Magis W, DeWitt MA, Wyman SK, et al. High-Level Correction of the Sickle Mutation is Amplified in Vivo During Erythroid Differentiation. Rochester, NY: Social Science Research Network; 2021.
135. Graphite Bio, Inc. A Phase I/II Study of GPH101 in Autologous CD34+ Hematopoietic Stem Cells to Convert HbS to HbA for Treating Severe Sickle Cell Disease. clinicaltrials.gov; 2022.
136. Blackwell RQ, Oemijati S, Pribadi W, Weng MI, Liu CS. Hemoglobin G Makassar: beta-6 Glu leads to Ala. *Biochim. Biophys. Acta*. 1970;214(3):396–401.
137. Viprakasit V, Wiriyasateinkul A, Sattayasevana B, Miles KL, Laosombat V. Hb G-Makassar [beta6(A3)Glu-->Ala; codon 6 (GAG-->GCG)]: molecular characterization, clinical, and hematological effects. *Hemoglobin*. 2002;26(3):245–253.
138. Sangkitporn S, Rerkamnuaychoke B, Sangkitporn S, Mitrakul C, Sutivigit Y. Hb G Makassar (beta 6:Glu-Ala) in a Thai family. *J. Med. Assoc. Thail. Chotmaihet Thangphaet*. 2002;85(5):577–582.
139. Mohamad AS, Hamzah R, Selvaratnam V, Yegapan S, Sathar J. Human hemoglobin G-Makassar variant masquerading as sickle cell anemia. *Hematol. Rep.* 2018;10(3):7210.
140. Mayuranathan T. Adenosine Base Editing of γ-Globin Promoters Induces Fetal Hemoglobin and Inhibit Erythroid Sickling. 2020;
141. Menzel S, Garner C, Gut I, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat. Genet.* 2007;39(10):1197–1199.
142. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105(5):1620–1625.
143. Luc S, Huang J, McEldoon JL, et al. Bcl11a-deficiency leads to hematopoietic stem cell defects with an aging-like phenotype. *Cell Rep.* 2016;16(12):3181–3194.
144. Wu Y, Zeng J, Roscoe BP, et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat. Med.* 2019;25(5):776–783.
145. Demirci S, Zeng J, Wu Y, et al. BCL11A enhancer-edited hematopoietic stem cells persist in rhesus monkeys without toxicity. *J. Clin. Invest.* 2020;130(12):6677–6687.
146. Psatha N, Reik A, Phelps S, et al. Disruption of the BCL11A Erythroid Enhancer Reactivates Fetal Hemoglobin in Erythroid Cells of Patients with β-Thalassemia Major. *Mol. Ther. Methods Clin. Dev.* 2018;10:313–326.
147. Lessard S, Rimmele P, Ling H, et al. Zinc Finger Nuclease-Mediated Disruption of the BCL11A Erythroid Enhancer Results in Enriched Biallelic Editing, Increased Fetal Hemoglobin, and Reduced Sickling in Erythroid Cells Derived from Sickle Cell Disease Patients. *Blood*. 2019;134:974.
148. Alavi A, Krishnamurti L, Abedi M, et al. Preliminary Safety and Efficacy Results from Precizn-1: An Ongoing Phase 1/2 Study on Zinc Finger Nuclease-Modified Autologous CD34+ HSPCs for Sickle Cell Disease (SCD). *Blood*. 2021;138(Supplement 1):2930–2930.

149. Walters MC, Smith AR, Schiller GJ, et al. Updated Results of a Phase 1/2 Clinical Study of Zinc Finger Nuclease-Mediated Editing of Poster 3974 BCL11A in Autologous Hematopoietic Stem Cells for Transfusion-Dependent Beta Thalassemia. 1.
150. Intellia Therapeutics Announces New, Robust Genome Editing Data for Sickle Cell Disease at the American Society of Hematology Meeting - Intellia Therapeutics.
151. Métais J-Y, Doerfler PA, Mayuranathan T, et al. Genome editing of HBG1 and HBG2 to induce fetal hemoglobin. *Blood Adv.* 2019;3(21):3379–3392.
152. Traxler EA, Yao Y, Wang Y-D, et al. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat. Med.* 2016;22(9):987–990.
153. Humbert O, Radtke S, Samuelson C, et al. Therapeutically relevant engraftment of a CRISPR-Cas9–edited HSC-enriched population with HbF reactivation in nonhuman primates. *Sci. Transl. Med.* 2019;11(503):.
154. Heath J, de Dreuzy E, Sousa P, et al. EDIT-301: AN AUTOLOGOUS CELL THERAPY TO PROMOTE FETAL HEMOGLOBIN EXPRESSION FOR THE POTENTIAL TREATMENT OF SICKLE CELL DISEASE. 2020;13.
155. Dreuzy ED, Heath J, Sousa P, et al. PRECLINICAL DATA SUPPORTING THE INITIATION OF THE EDIT-301 PHASE I/II RUBY CLINICAL TRIAL FOR THE POTENTIAL TREATMENT OF SICKLE CELL DISEASE. 2021;13.
156. Weber L, Frati G, Felix T, et al. Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. *Sci. Adv.* 2020;6(7):.
157. Mingoia M, Caria CA, Ye L, et al. Induction of therapeutic levels of HbF in genome-edited primary β0 39-thalassaemia haematopoietic stem and progenitor cells. *Br. J. Haematol.* 2021;192(2):395–404.
158. Li C, Psatha N, Sova P, et al. Reactivation of γ-globin in adult β-YAC mice after ex vivo and in vivo hematopoietic stem cell genome editing. *Blood*. 2018;131(26):2915–2928.
159. Gaudelli NM, Lam DK, Rees HA, et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* 2020;38(7):892–900.
160. Wang L, Li L, Ma Y, et al. Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. *Cell Res.* 2020;30(3):276–278.
161. Koblan LW, Doman JL, Wilson C, et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* 2018;36(9):843–846.
162. Frati G, Antoniou P, Hardouin G, Mlaya B, Miccio A. Editing the LRF Repressor Binding Site in the γ-Globin Promoters Induces Therapeutically Relevant Fetal Hemoglobin Levels for the Treatment of β-Hemoglobinopathies. 2020;
163. Thein SL. β-Thalassaemia. *Baillières Clin. Haematol.* 1998;11(1):91–126.
164. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull. World Health Organ.* 2008;86(6):480–487.
165. Thein SL. The Molecular Basis of β-Thalassemia. *Cold Spring Harb. Perspect. Med.* 2013;3(5):a011700.
166. Cao A, Gossens M, Pirastu M. Beta thalassaemia mutations in Mediterranean populations. *Br. J. Haematol.* 1989;71(3):309–312.
167. Ithanet.eu. IthaGenes, IthaID: 113. *IthaGenes IthaID 113*. 2022;
168. Busslinger M, Moschonas N, Flavell RA. Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell*. 1981;27(2 Pt 1):289–298.
169. Patsali P, Papasavva P, Stephanou C, et al. Short-hairpin RNA against aberrant HBBIVSI-110(G>A) mRNA restores β-globin levels in a novel cell model and acts as mono- and combination therapy for β-thalassemia in primary hematopoietic stem cells. *Haematologica*. 2018;103(9):e419–e423.

170. Xu S, Luk K, Yao Q, et al. Editing aberrant splice sites efficiently restores β-globin expression in β-thalassemia. *Blood*. 2019;133(21):2255–2262.
171. Patsali P, Turchiano G, Papasavva P, et al. Correction of IVS I-110(G>A) β-thalassemia by CRISPR/Cas-and TALEN-mediated disruption of aberrant regulatory elements in human hematopoietic stem and progenitor cells. *Haematologica*. 2019;104(11):e497–e501.
172. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 2018;24(7):927–930.
173. Ramadier S, Chalumeau A, Felix T, et al. Combination of lentiviral and genome editing technologies for the treatment of sickle cell disease. *Mol. Ther. J. Am. Soc. Gene Ther.* 2022;30(1):145–163.
174. Ren Q, Sretenovic S, Liu S, et al. PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat. Plants*. 2021;7(1):25–33.
175. Kurt IC, Zhou R, Iyer S, et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* 2021;39(1):41–46.
176. Musallam KM, Rivella S, Vichinsky E, Rachmilewitz EA. Non-transfusion-dependent thalassemias. *Haematologica*. 2013;98(6):833–844.
177. Steinberg MH. Fetal hemoglobin in β hemoglobinopathies: Is enough too much? *Am. J. Hematol.* 2022;
178. Topfer S, Feng R, Huang P, et al. Disrupting the adult-globin promoter alleviates promoter competition and reactivates foetal-globin gene expression. *Blood*. 2022;blood.2021014205.
179. Giarratana M-C, Kobari L, Lapillonne H, et al. Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat. Biotechnol.* 2005;23(1):69–74.
180. Kluesner MG, Nedveck DA, Lahr WS, et al. EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.* 2018;1:239–250.
181. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 2014;42(22):e168.
182. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol. Ther. Nucleic Acids*. 2014;3(12):e214.
183. Clement K, Rees H, Canver MC, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 2019;37(3):224–226.
184. Taher AT, Weatherall DJ, Cappellini MD. Thalassaemia. *Lancet Lond. Engl.* 2018;391(10116):155–167.
185. Ithanet.eu. IthaGenes, IthaID: 142. *IthaGenes IthaID 142*. 2022;
186. Gorski J, Fiori M, Mach B. A new nonsense mutation as the molecular basis for β° thalassaemia. *J. Mol. Biol.* 1982;154(3):537–540.
187. Ithanet.eu. IthaGenes, IthaID: 200. *IthaGenes IthaID 200*. 2022;
188. Treisman R, Proudfoot NJ, Shander M, Maniatis T. A single-base change at a splice site in a β0-thalassemic gene causes abnormal RNA splicing. *Cell*. 1982;29(3):903–911.
189. Chin JY, Kuan JY, Lonkar PS, et al. Correction of a splice-site mutation in the beta-globin gene stimulated by triplex-forming peptide nucleic acids. *Proc. Natl. Acad. Sci.* 2008;105(36):13514–13519.
190. Cosenza LC, Gasparello J, Romanini N, et al. Efficient CRISPR-Cas9-based genome editing of β-globin gene on erythroid cells from homozygous β039-thalassemia patients. *Mol. Ther. - Methods Clin. Dev.* 2021;21:507–523.
191. Milyavsky M, Gan OI, Trottier M, et al. A Distinctive DNA Damage Response in Human Hematopoietic Stem Cells Reveals an Apoptosis-Independent Role for p53 in Self-Renewal. *Cell Stem Cell*. 2010;7(2):186–197.
192. Cromer MK, Vaidyanathan S, Ryan DE, et al. Global Transcriptional Response to

CRISPR/Cas9-AAV6-Based Genome Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther. J. Am. Soc. Gene Ther.* 2018;26(10):2431–2442.
193. Schiroli G, Conti A, Ferrari S, et al. Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell*. 2019;24(4):551-565.e8.
194. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 2018;36(8):765–771.
195. Boutin J, Rosier J, Cappellen D, et al. CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. *Nat. Commun.* 2021;12(1):4922.
196. Leibowitz ML, Papathanasiou S, Doerfler PA, et al. Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.* 2021;53(6):895–905.
197. Turchiano G, Andrieux G, Klermund J, et al. Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell*. 2021;28(6):1136-1147.e5.
198. Orkin SH, Kazazian HH, Antonarakis SE, et al. Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. *Nature*. 1982;296(5858):627–631.
199. Derakhshan S, A K. Spectrum of β-Globin Gene Mutations and β-Thalassemia Haplotype Analysis among the Iranian Azeri Turkish Population. *Epidemiol. Open Access*. 2015;05:
200. Hoyer JD, McCormick DJ, Snow K, et al. FOUR NEW β CHAIN HEMOGLOBIN VARIANTS WITHOUT CLINICAL OR HEMATOLOGICAL EFFECTS: Hb SAN BRUNO [β39(C5)Gln→His]; Hb FORT DODGE [β93(F9)Cys→Tyr]; Hb RHODE ISLAND [β116(G18)His→Tyr]; AND Hb INGLEWOOD [β142(H20)Ala→Thr]. *Hemoglobin*. 2002;26(3):299–303.
201. Brown WJ, Niazi GA, Jayalakshmi M, Abraham EC, Huisman THJ. Hemoglobin athens-georgia, or α2β240(C6)Arg→Lys, a hemoglobin variant with an increased oxygen affinity. *Biochim. Biophys. Acta BBA - Protein Struct.* 1976;439(1):70–76.
202. Moo-Penn WF, Johnson MH, Bechtel KC, et al. Hemoglobins Austin and Waco: Two Hemoglobins with substitutions in the α1β2 contact region. *Arch. Biochem. Biophys.* 1977;179(1):86–94.
203. Henderson SJ, Timbs AT, McCarthy J, et al. Ten Years of Routine **α** - and **β** -Globin Gene Sequencing in UK Hemoglobinopathy Referrals Reveals 60 Novel Mutations. *Hemoglobin*. 2016;40(2):75–84.
204. Stabler SP, Jones RT, Head C, Shih DT, Fairbanks VF. Hemoglobin Denver [alpha 2 beta 2(41) (C7) Phe-->Ser]: a low-O2-affinity variant associated with chronic cyanosis and anemia. *Mayo Clin. Proc.* 1994;69(3):237–243.
205. Fomenko A, Kolokowski T, Heyse D, et al. Hemoglobin Rothschild – Unimpaired physical performance and oxygen uptake – A case report with literature review. *Respir. Med. Case Rep.* 2022;38:101681.
206. Renoux C, Feray C, Joly P, Lacan P, Francina A. Description of Three New α Variants and Four New β Variants: Hb Montluel [α110(G17)Ala → Val; HBA1: c.332C > T], Hb Cap d'Agde [α131(H14)Ser → Cys; HBA2: c.395C > G] and Hb Corsica [α100(G7)Leu → Pro; HBA1: 302T > C]; Hb Nîmes [β104(G6)Arg → Gly; HBB: c.313A > G], Hb Saint Marcellin [β112(G14)Cys → Gly; HBB: c.337T > G], Hb Saint Chamond [β80(EF4)Asn → 0; HBB: c.241_243delAAC] and Hb Dompierre [β29(B11)Gly → Arg; HBB: c.88G > C]. *Hemoglobin*. 2015;39(3):147–151.
207. rs1468286413 RefSNP Report - dbSNP - NCBI.

208. Inácio Â, Silva AL, Pinto J, et al. Nonsense Mutations in Close Proximity to the Initiation Codon Fail to Trigger Full Nonsense-mediated mRNA Decay*. *J. Biol. Chem.* 2004;279(31):32170–32180.
209. Psatha N, Georgakopoulou A, Li C, et al. Enhanced HbF reactivation by multiplex mutagenesis of thalassemic CD34+ cells in vitro and in vivo. *Blood*. 2021;blood.2020010020.
210. Clement K, Rees H, Canver MC, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 2019;37(3):224–226.
211. Alateeq S, Ovchinnikov D, Tracey T, et al. Identification of on-target mutagenesis during correction of a beta-thalassemia splice mutation in iPS cells with optimised CRISPR/Cas9-double nickase reveals potential safety concerns. *APL Bioeng.* 2018;2(4):046103.
212. Piel FB, Hay SI, Gupta S, Weatherall DJ, Williams TN. Global Burden of Sickle Cell Anaemia in Children under Five, 2010–2050: Modelling Based on Demographics, Excess Mortality, and Interventions. *PLOS Med.* 2013;10(7):e1001484.
213. Magrin E, Semeraro M, Hebert N, et al. Long-term outcomes of lentiviral gene therapy for the β-hemoglobinopathies: the HGB-205 trial. *Nat. Med.* 2022;28(1):81–88.
214. Chu SH, Packer M, Rees H, et al. Rationally Designed Base Editors for Precise Editing of the Sickle Cell Disease Mutation. *CRISPR J.* 2021;4(2):169–177.
215. Banno S, Nishida K, Arazoe T, Mitsunobu H, Kondo A. Deaminase-mediated multiplex genome editing in Escherichia coli. *Nat. Microbiol.* 2018;3(4):423–429.
216. Ryu S-M, Koo T, Kim K, et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* 2018;36(6):536–539.
217. Katti A, Dow LE. Base editing goes into hyperdrive. *Nat. Cell Biol.* 2020;22(6):617–618.
218. HbVar ID 2753.
219. Colak A, Toprak B, Yararbas K, Akyol F, Ceylan C. First Observation of Hemoglobin Jabalpur [Beta 3 (NA3) Leu> Pro] in the Turkish Population. *Turk J Hematol*. 2014;31:.
220. Romero Z, Lomova A, Said S, et al. Editing the Sickle Cell Disease Mutation in Human Hematopoietic Stem Cells: Comparison of Endonucleases and Homologous Donor Templates. *Mol. Ther. J. Am. Soc. Gene Ther.* 2019;27(8):1389–1406.
221. Hoban MD, Cost GJ, Mendel MC, et al. Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood*. 2015;125(17):2597–2604.
222. Tsai SQ, Zheng Z, Nguyen NT, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 2015;33(2):187–197.
223. Tan J, Zeng D, Zhao Y, et al. PhieABEs: a PAM-less/free high-efficiency adenine base editor toolbox with wide target scope in plants. *Plant Biotechnol. J.* 2022;20(5):934–943.
224. Noronha PA, Vida LN, Park CL, Honig GR. Hemoglobin-Specific Antibody in a Multiply Transfused Patient With Sickle Cell Disease. *Blood*. 1997;89(6):2155–2158.
225. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019;576(7785):149–157.
226. Petri K, Zhang W, Ma J, et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* 2021;1–5.
227. Kurita R, Suda N, Sudo K, et al. Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLOS ONE*. 2013;8(3):e59890.
228. Canver MC, Smith EC, Sher F, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015;527(7577):192–197.
229. Tanhehco YC. Gene therapy for hemoglobinopathies. *Transfus. Apher. Sci. Off. J. World Apher. Assoc. Off. J. Eur. Soc. Haemapheresis*. 2021;60(1):103061.
230. Magrin E, Miccio A, Cavazzana M. Lentiviral and genome-editing strategies for the

treatment of β-hemoglobinopathies. *Blood*. 2019;134(15):1203–1213.
231. Wienert B, Martyn GE, Funnell APW, Quinlan KGR, Crossley M. Wake-up Sleepy Gene: Reactivating Fetal Globin for β-Hemoglobinopathies. *Trends Genet.* 2018;34(12):927–940.
232. Mettananda S. Genetic and Epigenetic Therapies for β-Thalassaemia by Altering the Expression of α-Globin Gene. *Front. Genome Ed.* 2021;3:.
233. Cromer MK, Camarena J, Martin RM, et al. Gene replacement of α-globin with β-globin restores hemoglobin balance in β-thalassemia-derived hematopoietic stem and progenitor cells. *Nat. Med.* 2021;27(4):677–687.
234. Pavani G, Fabiano A, Laurent M, et al. Correction of β-thalassemia by CRISPR/Cas9 editing of the α-globin locus in human hematopoietic stem cells. *Blood Adv.* 2021;5(5):1137–1153.
235. Eaton WA, Bunn HF. Treating sickle cell disease by targeting HbS polymerization. *Blood*. 2017;129(20):2719–2726.
236. Badat M, Hua P, Mettananda S, et al. P104: BASE EDITING REPAIRS THE HBE MUTATION RESTORING THE PRODUCTION OF NORMAL GLOBIN CHAINS IN SEVERE HBE/β-THALASSEMIA PATIENT HEMATOPOIETIC STEM AND ERYTHROID CELLS. *HemaSphere*. 2022;6:19–20.
237. IthaGenes: an interactive database for haemoglobin variations and epidemiology - PubMed.
238. Miccio A, Cesari R, Lotti F, et al. In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of β-thalassemia. *Proc. Natl. Acad. Sci.* 2008;105(30):10547–10552.
239. Andreani M, Testi M, Gaziev J, et al. Quantitatively different red cell/nucleated cell chimerism in patients with long-term, persistent hematopoietic mixed chimerism after bone marrow transplantation for thalassemia major or sickle cell disease. *Haematologica*. 2011;96(1):128–133.
240. Magnani A, Pondarré C, Bouazza N, et al. Extensive multilineage analysis in patients with mixed chimerism after allogeneic transplantation for sickle cell disease: insight into hematopoiesis and engraftment thresholds for gene therapy. *Haematologica*. 2020;105(5):1240–1247.
241. El Hoss S, El Nemer W, Rees DC. Precision Medicine and Sickle Cell Disease. *HemaSphere*. 2022;6(9):e762.
242. bluebird bio Reports Second Quarter Financial Results and Provides Operational Update - bluebird bio, Inc.
243. What Bluebird Bio Gets Wrong In Pricing For Its $1.8 Million Drug.
244. bluebird bio Announces U.S. Commercial Infrastructure to Enable Patient Access to ZYNTEGLO®, the First and Only FDA-Approved Gene Therapy for People with Beta-Thalassemia Who Require Regular Red Blood Cell Transfusions - bluebird bio, Inc.
245. Johnson KM, Jiao B, Ramsey SD, et al. Lifetime medical costs attributable to sickle cell disease among nonelderly individuals with commercial insurance. *Blood Adv.* 2022;bloodadvances.2021006281.
246. Abou-El-Enein M, Angelis A, Appelbaum FR, et al. Evidence generation and reproducibility in cell and gene therapy research: A call to action. *Mol. Ther. Methods Clin. Dev.* 2021;22:11–14.
247. de la Fuente J, Gluckman E, Makani J, et al. The role of haematopoietic stem cell transplantation for sickle cell disease in the era of targeted disease-modifying therapies and gene editing. *Lancet Haematol.* 2020;7(12):e902–e911.
248. Coquerelle S, Ghardallou M, Rais S, et al. Innovative Curative Treatment of Beta Thalassemia: Cost-Efficacy Analysis of Gene Therapy Versus Allogenic Hematopoietic

Stem-Cell Transplantation. *Hum. Gene Ther.* 2019;30(6):753–761.
249. Rees HA, Minella AC, Burnett CA, Komor AC, Gaudelli NM. CRISPR-derived genome editing therapies: Progress from bench to bedside. *Mol. Ther. J. Am. Soc. Gene Ther.* 2021;29(11):3125–3139.
250. Masiuk KE, Brown D, Laborada J, et al. Improving Gene Therapy Efficiency through the Enrichment of Human Hematopoietic Stem Cells. *Mol. Ther.* 2017;25(9):2163–2175.
251. Hsieh MM, Bonner M, Pierciey FJ, et al. Myelodysplastic syndrome unrelated to lentiviral vector in a patient treated with gene therapy for sickle cell disease. *Blood Adv.* 2020;4(9):2058–2063.
252. Lidonnici MR, Ferrari G. Gene therapy and gene editing strategies for hemoglobinopathies. *Blood Cells. Mol. Dis.* 2018;70:87–101.
253. Li C, Georgakopoulou A, Newby GA, et al. In vivo base editing by a single intravenous vector injection for treatment of hemoglobinopathies. *JCI Insight*. 2022;
254. Koniali L, Lederer CW, Kleanthous M. Therapy Development by Genome Editing of Hematopoietic Stem Cells. *Cells*. 2021;10(6):1492.
255. Cortabarria AS de V, Makhoul L, Strouboulis J, et al. In utero Therapy for the Treatment of Sickle Cell Disease: Taking Advantage of the Fetal Immune System. *Front. Cell Dev. Biol.* 2021;8:.
256. Tai-MacArthur S, Lombardi G, Shangaris P. The Theoretical Basis of In Utero Hematopoietic Stem Cell Transplantation and Its Use in the Treatment of Blood Disorders. *Stem Cells Dev.* 2021;30(2):49–58.
257. Shangaris P, Loukogeorgakis SP, Subramaniam S, et al. In Utero Gene Therapy (IUGT) Using GLOBE Lentiviral Vector Phenotypically Corrects the Heterozygous Humanised Mouse Model and Its Progress Can Be Monitored Using MRI Techniques. *Sci. Rep.* 2019;9(1):11592.

Annex 1: Base-editing-mediated dissection of a γ-globin cis-regulatory element for the therapeutic reactivation of fetal hemoglobin expression

Base-editing-mediated dissection of a γ-globin *cis*-regulatory element for the therapeutic reactivation of fetal hemoglobin expression

Panagiotis Antoniou[1], Giulia Hardouin[1,2,3], Pierre Martinucci[1], Giacomo Frati[1], Tristan Felix[1], Anne Chalumeau[1], Letizia Fontana[1], Jeanne Martin[1], Cecile Masson[4], Megane Brusson[1], Giulia Maule[5], Marion Rosello[6], Carine Giovannangeli[7], Vincent Abramowski[8], Jean-Pierre de Villartay[8], Jean-Paul Concordet[7], Filippo Del Bene[6], Wassim El Nemer[9,10], Mario Amendola[11,12], Marina Cavazzana[3,13,14], Anna Cereseto[5], Oriana Romano[15], Annarita Miccio[1]*

[1]Université Paris Cité, Imagine Institute, Laboratory of chromatin and gene regulation during development, INSERM UMR 1163, 75015, Paris, France
[2]Université Paris Cité, Imagine Institute, Laboratory of Human Lymphohematopoiesis, INSERM UMR 1163, 75015, Paris, France
[3]Biotherapy Department and Clinical Investigation Center, Assistance Publique Hopitaux de Paris, INSERM, 75015 Paris, France
[4]Bioinformatics Platform, Imagine Institute, 75015, Paris, France.
[5]CIBIO, University of Trento, 38100, Trento, Italy
[6]Sorbonne Université, INSERM, CNRS, Institut de la Vision, 75015, Paris, France
[7]INSERM U1154, CNRS UMR7196, Museum National d'Histoire Naturelle, Paris, France
[8]Université Paris Cité, Imagine Institute, Laboratory of genome dynamics in the immune system, INSERM UMR 1163, 75015, Paris, France
[9]Établissement Français du Sang, UMR 7268, 13005, Marseille, France
[10]Laboratoire d'Excellence GR-Ex, 75015, Paris, France
[11]Genethon, 91000, Evry, France
[12]Université Paris-Saclay, Univ Evry, Inserm, Genethon, Integrare research unit UMR_S951, 91000, Evry, France
[13]Université Paris Cité, 75015, Paris, France
[14]Imagine Institute, 75015, Paris, France
[15] Department of Life Sciences, University of Modena and Reggio Emilia, 41125, Modena, Italy

*To whom correspondence should be addressed: A.M., Imagine Institute, 24, Boulevard du Montparnasse, 75015 Paris, France. E-mail address: annarita.miccio@institutimagine.org.

Keywords: base editing, hemoglobinopathies, fetal hemoglobin reactivation, hereditary persistence of fetal hemoglobin

Abstract

Sickle cell disease and β-thalassemia affect the production of the adult β-hemoglobin chain. The clinical severity is lessened by mutations that cause fetal γ-globin expression in adult life (i.e. the hereditary persistence of fetal hemoglobin). Mutations clustering ~200 nucleotides upstream of the *HBG* transcriptional start sites either reduce binding of the LRF repressor or recruit the KLF1 activator. Here, we use base editing to generate a variety of mutations in the -200 region of the *HBG* promoters, including potent combinations of four to eight γ-globin-inducing mutations. Editing of patient hematopoietic stem/progenitor cells is safe, leads to fetal hemoglobin reactivation and rescues the pathological phenotype. Creation of a KLF1 activator binding site is the most potent strategy – even in long-term repopulating hematopoietic stem/progenitor cells. Compared with a Cas9-nuclease approach, base editing avoids the generation of insertions, deletions and large genomic rearrangements and results in higher γ-globin levels. Our results demonstrate that base editing of *HBG* promoters is a safe, universal strategy for treating β-hemoglobinopathies.

Introduction

Sickle cell disease (SCD) and β-thalassemia are both genetic diseases caused by mutations in the β-globin locus. In SCD, a point mutation in the *HBB* gene leads to the formation of the sickle β^S-globin chain, which causes the polymerization of sickle hemoglobin (HbS), red blood cell (RBC) sickling, anemia, and organ damage[1,2]. In β-thalassemia, the partial or total absence of β-globin chains (β^+ and β^0, respectively) leads to the precipitation of noncoupled α-globin chains, apoptosis of erythroid precursors, ineffective erythropoiesis, and anemia[3–5]. Gene therapy approaches based on the transplantation of autologous, genetically modified hematopoietic stem cells (HSCs) have been investigated as a treatment option for patients lacking a compatible donor for allogeneic HSC transplantation[6].

The severity of both SCD and β-thalassemia is lessened by the hereditary persistence of fetal hemoglobin (HbF) in adulthood (HPFH)[7]. This persistence is due to mutations located 200 to 115 nucleotides upstream of the transcription start sites of the identical *HBG1* and *HBG2* γ-globin promoters. HPFH mutations either generate *de novo* DNA motifs recognized by transcriptional activators (e.g., KLF1)[8–10] or disrupt binding sites (BS) for transcriptional repressors (e.g., LRF and BCL11A)[11]. A comparison of HPFH mutations in the -200 region to identify the nucleotide change associated with the highest level of HbF expression (e.g., mutations disrupting the LRF BS vs mutation creating the KLF1 BS) has never been carried out. CRISPR-Cas9 nuclease strategies have been used to disrupt the LRF and BCL11A repressor BS via the non-homologous end-joining (NHEJ)-mediated generation of insertions/deletions (InDels) that mimic HPFH mutations and reactivate HbF expression[12–15]. Unfortunately, targeting the two identical γ-globin promoters can generate a 4.9-kb deletion that encompasses the *HBG2* gene and thus reduce overall HbF expression[16]. Furthermore, CRISPR-Cas9 nuclease cannot be used to introduce HPFH mutations creating an activator BS via homology-directed repair (HDR)[8], a pathway poorly active in human HSCs[17–19].

HSCs are highly sensitive to DNA double-strand breaks (DSBs)[20] – especially in case of multiple on-target events or concomitant on-target and off-target events. Even when highly specific single guide RNAs (sgRNAs) are used, the Cas9-sgRNA treatment of human HSPCs induces a DNA damage response (DDR) that can lead to apoptosis[21,22]. CRISPR-Cas9 can cause p53-dependent cell toxicity and cell cycle arrest, resulting in the selection of cells with a dysfunctional p53 pathway[23]. Furthermore, the generation of several on-target DSBs, simultaneous on-target and off-target DSBs, or even a single on-target DSB can lead to genomic deletions, inversions or translocations, chromosome loss, and chromothripsis[24–27]. Hence, the development of efficacious and safe treatment strategies for β-hemoglobinopathies based on precise base editing (rather than DSB-induced DNA repair) is highly desirable. Cytidine and adenine base editors (CBEs and ABEs) are composed of a Cas9 nickase and a deaminase that converts C-to-T and A-to-G[28], respectively. Base-editing approaches enable precise DNA repair with little or no DSB creation and thus rule out DSB-induced cytotoxicity and genomic rearrangements. Importantly, base editing results in homogeneous, predictable base changes, whereas NHEJ gives heterogeneous, unpredictable mutations. Recently, base-editing approaches targeting the *HBG* promoters showed some evidence of HbF reactivation in healthy donors' or β-thalassemic cells[29].

Here, we used CBEs and ABEs to dissect the -200 region of the *HBG1/2* promoters and to identify critical base conversions that induce changes in transcription factor occupancy (i.e., by creating a KLF1 activator BS and/or disrupting the LRF repressor BS) and lead to therapeutically relevant HbF levels.

Results

Generation of HPFH and HPFH-like mutations in erythroid cell lines

The majority of HPFH mutations in the -200 region of the *HBG* promoters reduce binding of the LRF repressor by disrupting its BS[30,31]. In the LRF BS, a total of 8 Cs can be converted to T by CBEs; this creates not only HPFH mutations but also additional HPFH-like mutations that might impair LRF binding (**Figure 1a**). In the same region, the -198 T>C HPFH mutation creates a *de novo* BS for the KLF1 activator[9] and probably disrupts the LRF BS[31] (**Figure 1a**). ABEs can be used to precisely

reproduce the -198 T>C HPFH mutation or to modify both the -198 and -199 central Ts in the LRF BS[28,32,33].

In K562 and HUDEP-2 erythroid cell lines, we identified the most efficient combinations of BEs and sgRNAs generating the following profiles: (i) LRF 8C (up to 8 Cs converted to Ts), using CBE-SpRY; (ii) LRF 4C (4 Cs converted to Ts), using CBE-SpRY; (iii) KLF1 (-198 T>C), using ABEmax, and (iv) LRF 2T (-198 and -199 T>C), using ABE8e (**Figure 1a, Supplementary Figures 1 and 2, and Supplementary note 1**). DSB-induced InDels were essentially absent (**Supplementary Figure 1e, Supplementary Figure 2d, Supplementary Note 2**), whereas the 4.9-kb deletion (resulting from the simultaneous cleavage of the identical *HBG1/2* promoters) was infrequent and occurred in ABEmax- and ABE8e-treated samples (**Supplementary Figure 1e and Supplementary Figure 2d**).

Chromatin immunoprecipitation (ChIP)-qPCR experiments in K562 cells showed that the LRF 8C profile and (to a lesser extent) the LRF 4C profile were associated with lower occupancy of the *HBG* promoters by LRF (**Figure 1b**). Disruption of the LRF BS (using ABE8e) also reduced LRF binding (LRF 2T**; Figure 1b**). Lastly, generation of a KLF1 BS impaired LRF binding - probably by altering the LRF binding motif and/or recruiting KLF1 and thus displacing LRF from the *HBG* promoters (**Figure 1b**). As expected, the frequency of base-edited *HBG* promoters was lower in LRF-immunoprecipitated samples than in the input DNA (**Figure 1b**). It is noteworthy that a base-editing frequency of ~50% was sufficient to reduce LRF binding to the same extent as in Cas9-nuclease-treated samples harboring >90% of edited alleles.

HbF reactivation after base editing of SCD HSPCs

In order to gauge the therapeutic potential of our base-editing strategies, we transfected primary human SCD HSPCs with plasmids encoding GFP and the BEs-sgRNA combinations that had performed best in K562 and HUDEP-2 cells. To compare our strategy with Cas9-nuclease-mediated disruption of the LRF BS, we used plasmids expressing a Cas9 nuclease and a sgRNA (-197) inducing InDels in the -200 region; these reduce LRF binding and reactivate HbF[12]. After transfection, GFP^{high} and GFP^{medium} cells were sorted (using FACS) to obtain populations with various editing efficiencies (**Figure 2a**).

GFP^{high} samples treated with CBEs (CBE-NCRH, CBE-SpG, or CBY-SpRY) and LRF_bs_3 sgRNA displayed an editing efficiency of 22.0%±2.6 (LRF 4C profile**; Figure 2b**). The LRF 8C profile was generated using CBE-SpRY and LRF_bs_2 sgRNA with an efficiency of 25.5%±3.6 (**Figure 2b**). GFP^{medium} populations transfected with CBEs were not edited. Samples carrying the LRF 2T (ABE8e-treated) profile or the KLF1 (ABEmax-treated) profile showed higher efficiencies in both GFP^{medium} and GFP^{high} bulk populations (respectively 56.5%±4.4 and 76.0%±1.4 for LRF 2T, and 41.0%±1.2 to 52.3%±2.7 for KLF1; **Figure 2b**). Sanger sequencing confirmed the absence of InDels, except in Cas9-treated cells (**Figure 2c**). The 4.9-kb deletion was detected in Cas9- and ABE8e-treated cells but occurred at a low frequency in the remaining samples (**Figure 2d**). NGS sequencing confirmed the Sanger sequencing data, and evidenced the simultaneous C>T conversions in samples carrying the LRF 4C and LRF 8C profiles, the simultaneous dual A>G conversion in the LRF 2T profile, and the precise creation of a KLF1 BS in the KLF1 profile (**Supplementary Figure 4a-d**). Cells treated with ABE8e mainly showed an LRF 2T profile, although a small proportion of the *HBG* promoters (11.3% of all base-editing events) carried a KLF1 BS. Similarly, in cells treated with ABEmax (KLF1 profile), a small proportion of *HBG* promoters harbored the LRF 2T profile (7.1% of all base-editing events) (**Supplementary Figure 4d**). As expected[12], Cas9 generated InDels in the LRF BS (**Supplementary Figure 4b-d**). Lastly, the NGS data confirmed the high level of product purity for all the BEs (**Supplementary Figure 4e**).

We next differentiated selected cultures from two SCD donors into RBCs. The erythroid differentiation was similar in the various groups, as measured by flow cytometry analysis of enucleated cells and early and late erythroid markers (**Supplementary Figure 5a-e**). CBE-treated samples bearing the LRF 4C or LRF 8C profile showed a low level of HbF reactivation (measured using CE-HPLC), as expected from the low observed base-editing efficiency (**Figure 2e**). ABE-treated samples (harboring either a KLF1 BS or the LRF 2T profile) expressed higher HbF levels. Although Cas9-treated samples presented the highest editing efficiency, they showed intermediate

HbF levels (**Figure 2e**). Similar results were observed for mRNA expression (using RT-qPCR) and single globin chain expression (using RP-HPLC) (**Supplementary Figure 6a-b**). Creation of a KLF1 BS was associated with high HbF expression – even at a low base-editing efficiency (39%; **Figure 2e**). Flow cytometry evidenced an elevated frequency of HbF-expressing cells in both the CBE- and ABE-treated samples. The frequency of HbS-expressing cells was lower in ABE-treated samples bearing the KLF1 or LRF 2T profile and (to a lesser extent) in Cas9-nuclease-treated samples compared to controls (**Figure 2f**).

To evaluate the effect of HbF reactivation on the sickling phenotype, we incubated RBCs under hypoxic conditions that induce HbS polymerization. After a 60-minute incubation, samples from donor 1 showed high proportions of non-sickle cells for all treated samples (52.5%, 58.4%, 75.5% and 68.2% for LRF 4C, LRF 8C, LRF 2T and Cas9-treated samples, respectively, versus 23.2% for the control sample); these values were in line with the corresponding HbF levels and base-editing efficiencies. Samples from donor 2 showed high proportions of corrected cells in the KLF1- and LRF 2T^{high}-bearing samples only (73.6%, 77.9%, 82.9% and 51.2% for KLF1^{med}, KLF1^{high}, LRF 2T^{high} and Cas9-treated samples, respectively, versus 14.7% for the control sample; **Figure 2g**). These results emphasized the high therapeutic potential of the ABE-mediated strategies and the need to exceed thresholds for base-editing efficiency and HbF expression when using CBEs to modify the Cs in the LRF BS.

HbF reactivation in single erythroid progenitors

Sorted GFP^{medium} and GFP^{high} SCD HSPCs were plated on a semi-solid medium that allows erythroid and granulocyte/monocyte differentiation at the clonal level (CFC assay). The numbers of erythroid (BFU-E) colonies and granulocyte/monocyte (CFU-GM) colonies were lower in plasmid-transfected samples than in mock-transfected samples (**Supplementary Figure 7a**). The base-editing efficiencies and InDel profiles in the BFU-E and CFU-GM pools were similar to those measured in liquid erythroid cultures (**Supplementary Figure 7b-d**). Significant γ-globin reactivation was observed in ABE-treated samples but only a mild increase in CBE-treated samples – probably because of the low editing efficiency (**Supplementary Figure 7e-f**). To accurately compare the efficacy of the various editing approaches, we measured γ-globin expression at the clonal level in BFU-Es. We observed a positive correlation between base-editing efficiency and γ-globin expression in all groups. Deviation of the base-editing efficiency from the expected 25% intervals (corresponding to one, two, three or four edited promoters) and the presence of BFU-E mosaics for *HBG* mutations indicated editing over several progenitor divisions, as reported previously[16]. Generation of a KLF1 BS was the most potent event for γ-globin reactivation (**Figure 2h**). LRF 8C and LRF 2T samples showed similar γ-globin levels, which were higher than those in LRF 4C samples (**Figure 2h**). As observed in liquid erythroid cultures, NGS analysis showed that in all the colonies treated with ABE8e, in a small fraction of *HBG* promoters only one of the 2 Ts was converted to C, thus generating a KLF1 BS (**Supplementary Figure 8**), which probably contributes to HbF reactivation. Almost all treated colonies (96.7%) showed no InDels (**Figure 2i**). The 4.9-kb deletion was detected in LRF 2T, KLF1 and LRF 4C colonies, although the frequency was relatively low (**Figure 2j**). As expected (given the DSB-free nature of base editing), we did not detect loss of the p arm of chromosome 11 (i.e. loss caused by DSBs at the *HBB* locus) (**Figure 2k**)[25].

RNA delivery of BEs in HSPCs rescues the sickling phenotype

With a view to developing a clinically relevant, selection-free system for delivering the base-editing machinery, we next sought to optimize an efficient, minimally toxic method for RNA transfection. To increase the base-editing efficiency of the CBEs (which were less efficient than ABEs in plasmid-transfected HSPCs, **Figure 2**), we optimized the plasmid encoding the CBE-SpRY-GFP fusion protein (capable of generating both 4C and 8C profiles) for *in vitro* transcription (**Supplementary Figure 9a**). We inserted a T7 promoter followed by a G (to allow efficient capping (Cap0) of the mRNA and to enhance translation), two copies of the 3' untranslated region (UTR) of the *HBB* gene (to increase the mRNA half-life and protein levels[34–36]), and a poly-A sequence after the 3' UTR (to further stabilize the mRNA)[37]. Lastly, we used uridine depletion to reduce the CBE mRNA's

immunogenicity. In K562 cells, the optimized construct (CBE-SpRY-OPT1) outperformed the original plasmid with regard to GFP expression and base-editing efficiency (**Supplementary Figure 9b-d**).

In vitro transcribed CBE-SpRY-OPT1, ABEmax and ABE8e mRNAs were transfected along with chemically modified sgRNAs into SCD HSPCs (**Figure 3a**). CBE-SpRY mRNA was co-transfected with LRF_bs_1 or LRF_bs_2 sgRNA to generate the LRF 4C or LRF 8C editing profiles, respectively. ABEmax and ABE8e coupled with KLF1_bs_1 sgRNA were used to generate the KLF1 and LRF 2T editing profiles, respectively. In parallel, we applied a delivery method currently used in the clinic[38], i.e. the transfection of a ribonucleoprotein (RNP) complex containing Cas9 and the -197 sgRNA[12].

In liquid erythroid cultures, we generated LRF 4C, LRF 8C, LRF 2T and KLF1 profiles with efficiencies of 36.7%±6.2, 45.3%±8.1, 49.7%±10.5 and 48.0%±2.9, respectively, on day 6 (**Figure 3b-e**). Much the same frequencies were observed on day 13, indicating that base-edited cells were not selected against during erythroid differentiation. Sanger sequencing revealed InDels in Cas9-treated samples only (79.3%±2.3) (**Figure 3f**). The frequency of the 4.9-kb deletion was negligible in RNA-transfected, base-edited samples (**Figure 3g**). Deep sequencing of the *HBG1/2* promoters confirmed the editing profiles observed with Sanger sequencing and the high product purity (**Supplementary Figure 10**).

The RNA-transfected and control samples did not differ significantly in the number of BFU-Es and CFU-GMs; this finding confirmed the safety of our RNA-based protocol with regard to progenitor viability (**Figure 3h**). Base-editing efficiencies tended to be higher in the BFU-E pools than in the CFU-GM pools (**Supplementary Figure 11a**). The Indel efficiency was negligible in base-edited colonies, and only some of the ABEmax-treated samples had the 4.9-kb deletion (**Supplementary Figure 11b-c**).

Next, we differentiated RNA-transfected SCD HSPCs into mature RBCs. The enucleation rate and the expression of erythroid markers were similar in control vs. edited cells (**Supplementary Figure 12a-e**). RT-qPCR, RP-HPLC and CE-HPLC measurements showed strong HbF reactivation and clinically relevant expression in all samples and especially in those carrying the KLF1 profile (71.3%±6.1) (**Figure 3i-k**). The LRF 4C, LRF 8C, LRF 2T, and Cas9-treated samples all had similar HbF levels (**Figure 3i-k**), although genome editing efficiency was lower in base-edited samples. Flow cytometry measurements confirmed the presence of HbF reactivation in all the edited samples. For LRF 2T and KLF1, we also observed a lower frequency of HbS-expressing cells (**Figure 3l-m**) compared to controls. Similar results were observed for pooled BFU-E colonies (**Supplementary Figure 11d-e**).

All the samples showed significantly lower frequencies of sickle cells (relative to controls) and thus produced enough HbF to inhibit Hb polymerization (**Figure 3n and Supplementary Figure 12f**). The generation of a KLF1 BS was even able to correct the sickling phenotype in the hard-to-correct RBCs from donor 3.

Base editing of the HBG promoters rescues the β-thalassemic phenotype

Using RNA transfection, we delivered base editors to HSPCs obtained from two β^0/β^+ β-thalassemia patients. Given the small number of available cells, we selected ABEmax/KLF1_bs_1 sgRNA to create the KLF1 profile (which had given the greatest HbF reactivation in SCD samples) and CBE-SpRY/LRF_bs_2 sgRNA to create the LRF 8C profile (which contained more HPFH- and HPFH-like mutations than LRF 4C and LRF 2T).

We observed efficiencies of 47.5%±4.5 and 69.0%±2.0 in LRF 8C and KLF1 liquid erythroid cultures, respectively, and only few InDels (**Figure 4a-c**). The total number of BFU-E and CFU-GM was similar in the edited populations and in control samples (**Figure 4d**). The editing efficiency in BFU-E was similar to that observed in the liquid erythroid cultures but was lower in CFU-GM (**Supplementary Figure 13a**). We observed low InDel and 4.9-kb deletion frequencies only in ABEmax-treated KLF1 BFU-E and CFU-GM pools (**Supplementary Figure 13b-c**). HbF reactivation was observed (using HPLC) in both LRF 8C and KLF1 BFU-E samples (**Supplementary Figure 13d-e**).

In liquid erythroid cultures, we observed strong γ-globin reactivation, (as assessed by RT-qPCR, HPLC, and flow cytometry) for both LRF 8C and KLF1 (**Figure 4e-h**). Levels of α-globin precipitates (a hallmark of β-thalassemia) were substantially lower in LRF 8C samples and were null in KLF1 samples (**Figure 4i**). The ratio between α and non-α globins reached normal levels (**Figure 4h**).

The delayed enucleation process observed in control samples (due to ineffective erythropoiesis) was rescued in the edited samples (**Figure 4j**). Moreover, enucleated cells were larger in treated samples than in control β-thalassemic cells (**Figure 4k**). The delayed erythroid differentiation typical of thalassemic cells had been corrected, as evaluated by the CD36 and CD71 expression profiles: at the end of the differentiation process, both of these erythroid markers were found to be correctly downregulated in samples derived from edited HSPCs (**Figure 4l-m**). The α4-integrin, GYPA and BAND3 erythroid markers had similar expression profiles during differentiation in all the samples (**Figure 4n-o**). In parallel, we observed a low frequency of apoptotic cells in samples carrying the LRF 8C or KLF1 profile, relative to controls (**Figure 4p**). The production of reactive oxygen species (ROS, as evaluated by flow cytometry) typically observed in β-thalassemic cells was much lower in KLF1 samples than in controls (**Figure 4q**).

Immune and DNA damage response in HSPCs

To minimize the immunogenicity of the CBE-SpRY mRNA, we exchanged the nucleotide that follows the T7 promoter (G>A); this allows alternative capping (Cap1) and reduces the likelihood of an innate immune response[39] (**Supplementary Figure 9a**). The CBE-SpRY-OPT2 plasmid was similar to CBE-SpRY-OPT1 with regard to GFP expression and base-editing efficiency in K562 cells (**Supplementary Figure 9b-d**). The CBE-SpRY-OPT2 plasmid was transcribed *in vitro* using Cap1 analog and 5-methoxyuridine (to completely eliminate uridines from the transcript[39,40] and further reduce the immunogenic potential) and purified using a silica membrane.

Mobilized healthy donor (HD) HSPCs were transfected with CBE-SpRY-OPT1 mRNA (uridine depleted and capped with a Cap0 analog), CBE-SpRY-OPT2 mRNA (uridine depleted, 5-methoxyuridine, and capped with a Cap1 analog), or unmodified ABEmax mRNA (non-uridine depleted and capped with a Cap0 analog). CBE-SpRY mRNA was co-transfected with the LRF_bs_2 sgRNA (LRF 8C profile) and ABEmax mRNA with the KLF1_bs_1 sgRNA (KLF1 profile). The base-editing efficiencies increased over time and reached maximum levels 6 days post-transfection (49.0%±3.0, 54.3%±3.5 and 47.3%±4.9 for CBE-SpRY-OPT1, CBE-SpRY-OPT2 and ABEmax, respectively; **Figure 5a-b**). In parallel, we used Cas9 RNP complexes that generate Indels in the -200 region. InDels were observed only in Cas9-treated samples, and 4.9-kb deletions only in ABEmax- and Cas9-treated samples (**Figure 5c-d**).

Firstly, we assessed the immune response to our RNAs by measuring the expression of genes activated by RNA stimuli. Twelve and 24 h after transfection, we did not detect any immune responses in either control or base-edited samples – even those treated with non-modified ABE mRNA constructs (**Supplementary Figure 14 and Figure 5e**). We then measured the expression of CDKN1 (p21) as a readout of p53-induced DDR. *CDKN1* was up-regulated in Cas9- and CBE-SpRY-treated HSPCs 48 h after treatment but not in ABEmax-treated samples (**Figure 5f**).

Base editing in HSPCs induces few transcriptomic changes

To examine the effect of base editing on the overall gene expression profile, we performed RNA-seq of control and edited HD HSPCs transfected with CBE-SpRY/LRF_bs_2 (LRF 8C profile), ABEmax/KLF1_bs_1 (KLF1 profile) or Cas9 RNP 48 h after transfection. Overall, we observed few differentially expressed genes (DEGs) (**Figure 6a**). There were 37 DEGs in Cas9-treated samples, 13 in CBE-SpRY-OPT2-treated cells bearing the LRF 8C profile, and 3 in KLF1 samples transfected with ABEmax (**Figure 6b**).

The genes dysregulated in CRISPR-Cas9-treated HSPCs (*CDKN1A*, *MIR34AHG*, *DDB2*, *ZMAT3*, *BAX*, *BBC3* and *RPS27L*) were involved in DDR and/or apoptosis. Genes upregulated in both Cas9- and CBE-treated samples included p53 targets (*CDKN1A* and *MIR34AHG)* and *PLXNB2,* which was shown to be involved in HSC self-renewal and proliferation in the mouse[41]*.* Other genes that were specifically upregulated in CBE-treated samples (*THBS1*, *PPM1F* and *VWF*)

have roles in apoptosis and HSC biology[42–45]. Downregulated genes in CBE-treated samples include *FN3K*, which protect cells from oxidative stress via NRF2 deglycation[46,47]. Interestingly, only globin genes were differentially expressed in ABEmax-treated cells: *HBG* upregulation was accompanied by a decreased synthesis of the adult *HBB* and *HBD* transcripts (**Figure 6b**).

To assess potential off-target activity of BEs, we further analyzed RNA-seq data from control and edited HD HSPCs. A similar number of C>T or A>G variants were observed in all the samples, independently of specific treatment; hence, no RNA deamination caused by CBE-SpRY or ABEmax expression was detected (**Figure 6c**).

DNA off-target activity of BEs

We assessed the off-target activity of all the sgRNAs via the genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq) in HEK293T cells. The Cas9-SpRY or Cas9 nuclease was combined with LRF_bs_3 sgRNA (LRF 4C profile), LRF_bs_2 sgRNA (LRF 8C profile), or KLF1_bs_1 sgRNA (KLF1 and the LRF 2T profiles). Only a few of the top 20 off-targets mapped to exons, and so a major impact on protein expression was unlikely (**Figure 7a-c**).

In erythroblasts derived from RNA-transfected SCD HSPCs, NGS of the top 10 sites and the off-targets located in exons did not reveal substantial DNA off-target activity in most cases (**Figure 7d-f**). Off-target activity was observed in an exon only for the KLF1_bs_1 sgRNA in LRF 2T samples (OT15-*CES4A*; 9.01%±1.8) but was present at background levels in KLF1 samples. We detected a low level of off-target activity in an intron in LRF 4C samples (OT1-*EHBP1*; 2.8%±0.5). KLF1_bs_1 sgRNA also showed some off-target activity at an intronic site; this activity was significantly lower for ABEmax than for ABE8e (OT4-*ITPR1*: 54.0%±3.1 in LRF 2T samples and 22.6%±2.3 in KLF1 samples). It is noteworthy that base editing at the OT4 site creates a SNP (rs760488983; T>C) that is reportedly not associated with any clinical manifestations. Lastly, we observed off-target activity in LRF 8C samples in an intergenic region (OT1: 48.6%±4.9) that, however, did not map to regulatory regions active in hematopoietic cells (**Supplementary Figure 15a**). Importantly, few or no indels were detected at the on-target sites or at the vast majority of off-target sites (**Supplementary Figure 15b-d**). The RNA-seq data showed that off-target activity did not affect the expression of targeted genes in HSPCs (**Figure 6b**).

Lastly, in CBE-SpRY-treated LRF 8C samples and ABEmax-treated KLF1 samples, whole exome sequencing (WES) analysis showed that for each donor, 99.9% of single nucleotide variants were shared between untreated and edited samples, indicating no detectable sgRNA-dependent or sgRNA-independent off-target activity in exons (**Figure 7g**).

Base editing of the HBG promoters in repopulating HSCs

To evaluate the ability of BEs to target the *HBG* promoters in repopulating HSCs, we xenotransplanted HD or SCD HSPCs transfected with CBE-SpRY-OPT2 mRNA and LRF_bs_2 sgRNA (LRF 8C) or with ABEmax mRNA and KLF1_bs_1 sgRNA (KLF1) into immunodeficient NBSGW mice (**Figure 8a**). Sixteen to 20 weeks post-transplantation, no differences were observed between edited and control HSPCs with regard to engraftment and differentiation potential, as measured by the frequency of human $CD45^+$ cells in the hematopoietic tissues and the expression of lineage-specific markers (**Figure 8b and Supplementary Figure 16**). Human $CD45^+$ bone marrow cells were isolated and subjected to a CFC assay. Mock and edited samples had similar clonogenic potentials, although the latter was slightly lower in CBE-treated cells (**Figure 8c**).

The base-editing efficiency in human bone marrow cells was 10.5%±5.4 for LRF 8C and 25.7%±6.6 for KLF1 (**Figure 8d**). Similar results were typically observed in spleen, blood and bone marrow BFU-E and CFU-GM pools (**Figure 8d**). Despite the lower efficiency, the editing profile of engrafted populations was similar to that of the input cells (**Figure 8d-e**). InDels were absent or scarce in engrafted cells (**Figure 8f**). The 4.9-kb deletion was detected in around 50% of the mice (**Figure 8g and Supplementary Figure 17**). It is noteworthy that in KLF1 samples, off-target activity at OT4 was detected in both input and engrafted populations (**Figure 8h**). The mean OT4 editing efficiency was similar in *in vivo* samples and input cells; however, the fact that value varied markedly from one sample to another *in vivo* (between 0 and 40%) ruled out a selective advantage of cells harboring off-target events (**Figure 8h)**

In KLF1 samples, 71.4% of bone-marrow-derived single BFU-E harbored at least one edited *HBG* promoter, while OT4 editing was observed in ~45% of colonies (mostly in those with edited *HBG* promoters) (**Figure 8i**). Chromosome arm loss was not detected in the edited colonies (**Figure 8j**). Lastly, BFU-E harboring a KLF1 BS showed elevated γ-globin expression levels, which were positively correlated with the number of edited promoters (**Figure 8k-l**)

Discussion

Clinical trials based on the lentivirus-mediated, permanent integration of a β-like globin gene or a BCL11A-downregulating microRNA have given promising results. However, lentivirus-based gene therapy only partially corrects the clinical phenotype of patients with severe β-thalassemia, and HbS levels still remain high in patients with SCD because the therapeutic transgene is not expressed sufficiently[48,49]. Furthermore, the use of these strategies is limited by the potential occurrence of insertional mutagenesis associated with lentiviral vectors.

Nuclease-based genome editing strategies have been developed for the treatment of β-hemoglobinopathies. Direct gene correction via HDR is poorly efficient in *bona fide* HSCs[50]. Attempts to reactivate HbF through the CRISPR-Cas9 nuclease-mediated disruption of the γ-globin genes' *cis*-regulatory elements (by mimicking HPFH mutations) have provided proof of efficacy in preclinical studies[12]. Nuclease-based reactivation of HbF through disruption of the γ-globin genes' *trans*-regulatory elements has already been tested in the clinic, with promising results[38]. However, safety concerns with regard to harmful DSBs and large genomic rearrangements in HSCs are still present[21–27].

Base editing has therefore been used to generate HPFH mutations[13,32] in HD cells or reduce the levels of the HbF repressor BCL11A[51] in cells from patients with β-hemoglobinopathies. Editing the BCL11A BS in the *HBG* promoters was moderately efficient, with frequencies of 7% to 22%[13,14]. In contrast (and despite a degree of cytotoxicity), two cycles of electroporation led to *BCL11A* editing rates of ~90% and therapeutically relevant HbF levels (~30%) in SCD RBCs; full correction of the β^+-thalassemic phenotype was achieved by combining *BCL11A* editing with correction of the disease-causing mutation[51]. Gaudelli et al. used ABEs to install the -198 T>C HPFH mutation or to simultaneously install the -198 and -199 mutations in HD HSPCs and provided proof of concept for HbF reactivation[32]. Nevertheless, the therapeutic potential and safety of these strategies had not been assessed in patient HSPCs. Furthermore, ABEs and CBEs targeting the -200 region of the *HBG* promoters and the specific HPFH or HPFH-like mutations induced in primary cells had not been compared.

In the present study, we used precise base editing to dissect the difficult-to-target LRF BS in the -200 region of the *HBG* promoters. Firstly, we screened sgRNAs and BEs in erythroid cell lines, in order to identify the most efficient combinations creating various editing profiles. In particular, we inserted a variety of HPFH and HPFH-like mutations creating the LRF 4C, LRF 8C, LRF 2T and KLF1 profiles. Interestingly, all the mutations (including the -198 T>C HPFH mutation that generates a KLF1 activator BS) led to weaker LRF binding. Editing both C stretches (LRF 8C profile) led to a more pronounced weakening in LRF binding and higher HbF levels, relative to the LRF 4C profile. This finding is line with the results of a recent study in which both C-stretches were important for LRF binding[31]. As the -198 or the -199 nucleotides contribute only moderately to LRF binding[31], the generation of a KLF1 activator BS might weaken LRF binding mainly through transcription factor competition, i.e., displacement of LRF by KLF1. Creation of a KLF1 BS gave the highest HbF levels; this finding indicates that removal of the LRF repressor is enough to reactivate *HBG* genes but that recruitment of an activator results in higher gene expression levels. Conversion of both -198 or -199 nucleotides (LRF 2T profile) also reduced LRF occupancy, suggesting that combinatorial editing can affect LRF binding. Furthermore, generation of a KLF1 BS in a fraction of *HBG* promoters in the LRF 2T samples might also have contributed to diminished LRF binding and to HbF reactivation.

Cas9-treated samples showed lower LRF binding, and the HbF levels were similar to those observed in LRF 2T, LRF 8C and LRF 4C samples. However, these levels were achieved with a frequency of edited promoters that was significantly higher than in base-edited samples. ^{G}γ-globin expression (due to loss of the *HBG2* gene) was lower in Cas9-treated samples than in base-edited samples (**Figure 3j**). Indeed, the frequency of the 4.9-kb deletion was substantially lower in base-

edited samples than in Cas9-treated samples. Even ABEmax-transfected cells (a small proportion of which had lost *HBG2* genes) did not show an imbalance between $^{G}\gamma$ and $^{A}\gamma$ chains (**Figure 3j**). Notably, Cas9-treated samples showed InDels mostly in the second stretch of Cs, thus affecting only nucleotides that are modified in the LRF 4C samples that gave the lowest level of HbF reactivation. Furthermore, some Cas9-induced InDels might have affected neighboring regulatory elements. According to a recent study, the -190 GATA motif (proximal to the -200 region) recruits activator when HPFH mutations are present, and GATA1 binding is essential for HbF reactivation[52]. The disruption of this motif in Cas9-treated samples might prevent the recruitment of GATA1. Indeed, deep sequencing of the *HBG1/2* promoters revealed the presence of deletions (albeit at low frequencies) affecting the -190 GATA motif. Overall, Cas9-mediated editing is therefore less predictable and less accurate than base editing, which allows the precise introduction of well-defined HPFH mutations in the -200 region of the *HBG* promoters.

Importantly, our base-editing strategies induced high levels of HbF expression and rescued both the SCD and β-thalassemic cell phenotypes. The thresholds of HbF expression needed to ameliorate the clinical manifestations of SCD have been defined as 70% of HbF-expressing cells and HbF accounting for 30% of the total quantity of hemoglobins[53,54]. *In vitro*, base-edited samples derived from SCD HSPCs and carrying the four editing profiles (with a frequency of 40-50%), exceeded this threshold; in particular, KLF1 samples expressing 70% of their hemoglobin as HbF. Accordingly, the pathological SCD cell phenotype was effectively reverted. Disruption of the LRF repressor BS or (more potently) the creation of a KLF1 activator BS reactivated HbF and rescued the β-thalassemic phenotype in erythroid cells differentiated from HSPCs collected from patients with β^0/β^+ thalassemia. The high HbF levels induced by generation of a KLF1 BS might be sufficient to correct the phenotype of β^0/β^0 thalassemic cells lacking residual β-globin expression.

The clinically approved methods of delivering genome editing tools to HSPCs include RNP or RNA transfection[55]. Unfortunately, we did not succeed in producing BE RNPs capable of editing the genome of HSPCs because of the low concentration and protein precipitation. Other researchers have recently reported that BE RNPs do not effectively edit primary HSPCs or T cells, as compared to BE mRNAs/sgRNA[56,57]. To develop a clinically relevant protocol for delivering the base-editing system to HSPCs[58], we therefore produced BE mRNAs and coupled them with chemically modified sgRNAs. This combination achieved efficient, selection-free base editing in SCD and β-thalassemic HSPCs; this was even true for CBEs that were inefficient with plasmid transfection. Our RNA-mediated base-editing protocol neither affected HSPC viability and differentiation nor induced an innate immune response in HSPCs.

Importantly, the near-total absence of indels mitigates concerns related to the targeting of four different genomic regions (e.g. a risk of large genomic rearrangements, due to multiple DSBs). However, low levels of the 4.9-kb deletion between the two identical *HBG1/2* promoters were detected in base-edited samples – primarily in samples transfected with ABE8e or ABEmax. It is noteworthy that these base editors generate 4.9-kb deletions but no InDels as we have shown previously for Cas9 nickase that generates precise editing with no DSB formation[59]. We hypothesize that the generation of two nicks in *cis* by an sgRNA induces the 4.9-kb deletion via strand displacement. This process might be facilitated by the homology of *HBG1* and *HBG2*. Importantly, deletions were occasionally observed in RNA-transfected samples and did not affect the *HBG1/2* globin ratio. Lastly, the persistence of ABEmax-induced deletions in vivo suggests that HSCs harboring a 4.9 kb deletion were not at a particular disadvantage.

Furthermore, we confirmed that Cas9 nuclease-mediated DSB formation in HSPCs is associated with activation of the p53 pathway. In contrast, our base-editing strategies (i.e. the LRF 8C and KLF1 profiles generated by CBE-SpRY and ABEmax, respectively) induced a lower or null DDR in HSPCs and thus offered a safer way to genetically manipulate HSPCs. HSPCs edited with ABEmax and control samples did not differ significantly with regard to the overall gene expression profile. Despite the absence of DSB-induced InDel formation, several genes in CBE-treated HSPCs were nevertheless dysregulated; some of these genes were involved in apoptosis and HSC biology. Accordingly, the base-editing efficiency was lower in repopulating HSCs treated with CBE-SpRY than in HSCs treated with ABEmax; hence, CBEs might have induced some toxicity *in vivo* or might be less efficient with *bona fide* HSCs. In contrast, we obtained evidence of (i) editing in a large

proportion of repopulating HSCs edited with ABEmax, and (ii) HbF reactivation in the HSCs' erythroid progeny. These results emphasize the importance of choosing the right base editor for HSC-based therapeutic strategies. In fact, to date, only ABE8e-NRCH[56] and A3A(N57)-BE3[51] (a CBE containing a different deaminase) have shown their effectiveness in human repopulating HSCs. The delivery of the CBEs and ABEs used in the present study to clinically relevant HSPCs could be further optimized with regard to the promoter editing frequency *in vitro* (e.g., in both BFU-E and CFU-GM), the base-editing efficiency in repopulating HSCs, and the fitness of the edited HSCs. By way of an example, we could use recently developed ABEs that are more efficient than ABEmax[32,60]. However, given the selective advantage of corrected erythroid precursors and RBCs, the creation of a KLF1 BS might be enough to strongly reactivate HbF and correct both SCD and β-thalassemia phenotypes *in vivo*.

Lastly, we comprehensively assessed the off-target activity of our base-editing systems in primary HSPCs. In particular, in the present study we have analyzed sgRNA-independent off-target RNA activity of CBE-SpRY and ABEmax in primary human HSPCs. Interestingly, transient mRNA delivery of CBEs and ABEs did not lead to greater deamination of the cellular transcriptome, as it is typically observed upon plasmid delivery[32,61]. Hence, the base editors used here appear to be safe in clinically relevant cells. Furthermore, the WES data did not reveal sgRNA-independent off-target activity within exons in CBE-SpRY-treated (LRF 8C) and ABEmax-treated (KLF1) samples. However, the limited WES coverage prevented us from identifying infrequent events. A few sgRNA-dependent DNA off-target edits in primary cells were observed for CBEs and ABEs, although there was no impact on gene expression and no occurrence of InDels potentially associated with large genomic rearrangements. The fact that off-target DNA activity was lower for ABEmax than for the highly processive and highly efficient ABE8e suggests that the former is the safer of the two [56]. The observation of ABEmax-mediated sgRNA-dependent off-target activity was observed also *in vivo,* suggests that it does not have a detrimental effect on HSC engraftment and differentiation. Furthermore, the use of a high-fidelity Cas9 nickase and PAM-restricted BEs (e.g., CBE-NCRH instead of CBE-SpRY) or the RNP-based delivery of the base-editing system might substantially reduce sgRNA-dependent DNA off-target activity[33,51,56,62].

In conclusion, our present results provided proof of concept for base-editing treatment strategies for both SCD and β-thalassemia. This universally applicable therapeutic strategy does not depend on the specific disease-causing mutation, and thus does not require the mutation-specific CRISPR-Cas9 tools, as described previously[51,56]. The translation of our approach into the clinic will require (i) optimization of genome editing in a sufficiently large number of HSCs that would allow effective reconstitution of the bone marrow and production of normal RBCs, and (ii) the establishment of a large-scale transfection protocol with clinical-grade reagents.

Methods

Cell line culture

The human fetal erythroleukemia cell line K562 was obtained commercially (ATCC). The human umbilical cord-derived erythroid progenitor HUDEP-2 cell line was obtained by the Cell Engineering Division of RIKEN BRC Cell Bank (Ibaraki, Japan)[63]. K562 cells were maintained in RPMI 1640 (Lonza) containing glutamine and supplemented with 10% fetal bovine serum (Lonza), 2 mM Hepes (Life Technologies), 100 nM sodium pyruvate (Life Technologies), and penicillin and streptomycin (Life Technologies). HUDEP-2 cells were cultured in StemSpan SFEM (Stem Cell Technologies), supplemented with 1 µg/mL doxycycline (Sigma), 10^{-6} M dexamethasone (Sigma), 100 ng/mL human stem cell factor (SCF) (Peprotech), 3 IU/mL erythropoietin (Necker Hospital Pharmacy), L-glutamine (Life Technologies) and penicillin/streptomycin.

HSPC purification and culture

We obtained human granulocyte colony-stimulating factor (G-CSF)-mobilized peripheral blood $CD34^+$ HSPCs from healthy donors, human non-mobilized and plerixafor-mobilized peripheral blood $CD34^+$ HSPCs from SCD patients and human non-mobilized peripheral blood $CD34^+$ HSPCs from β-thalassemia patients. SCD and β-thalassemic samples eligible for research purposes were obtained from the "Hôpital Necker-Enfants malades" Hospital (Paris, France). Healthy donors were

either obtained from the “Hôpital Necker-Enfants malades” Hospital (Paris, France) or purchased by Caltag. Written informed consent was obtained from all adult subjects. All experiments were performed in accordance with the Declaration of Helsinki. The study was approved by the regional investigational review board (reference: DC 2014-2272, CPP Ile-de-France II “Hôpital Necker-Enfants malades”). HSPCs were purified by immunomagnetic selection with AutoMACS (Miltenyi Biotec) after immunostaining with the CD34 MicroBead Kit (Miltenyi Biotec). Forty-eight hours before transfection, $CD34^+$ cells were thawed and cultured at a concentration of $5x10^5$ cells/ml in the “HSPC medium” containing StemSpan (STEMCELL Technologies) supplemented with penicillin/streptomycin (Gibco), 250 nM StemRegenin1 (STEMCELL Technologies), and the following recombinant human cytokines (PeproTech): human stem cell factor (SCF) (300 ng/ml), Flt-3L (300 ng/ml), thrombopoietin (TPO) (100 ng/ml), and interleukin-3 (IL-3) (60 ng/ml).

Plasmids

Plasmids used in this study include:
pCMV_ABEmax_P2A_GFP (Addgene #112101),
pCMV_AncBE4max_P2A_GFP (Addgene #112100),
SaKKH-ABEmax (Addgene #119815),
pBT374 (Addgene #125615),
pBT372 (Addgene #125613),
pMJ920 (Addgene #42234),
ABE8e (Addgene #138489),
pCMV-BE4max-NRCH (Addgene #136920),
pCMV-BE4max-NRRH (Addgene #136918),
pCAG-CBE4max-SpG-P2A-EGFP (RTW4552) (Addgene #139998)
pCAG-CBE4max-SpRY-P2A-EGFP (RTW5133) (Addgene #139999)
pCMV-T7-SpRY-P2A-EGFP (RTW4830) (Addgene #139989).
The AncBE4max_NAA plasmid was created by replacing the sequence encoding the PAM interaction domain of the SpCas9 nickase (SpCas9n) with the one of the SmaCas9[64]. The sequence encoding the SmaCas9 domain was codon optimized by using the Genscript codon optimization software and obtained by gene synthesis (Genscript). The SaKKH-AncBE4max plasmid was created by combining the sequences encoding the deaminase domain from the AncBE4max plasmid (Addgene #112094) and the SaKKH-Cas9 nickase from the SaKKH-ABEmax plasmid (Addgene #119815).
A DNA fragment (3’UTR+poly-A) containing two copies of the 3’ untranslated region (UTR) of the *HBB* gene and a poly-A sequence of 96 adenines was purchased by Genscript. Similarly, another DNA fragment containing the uridine-depleted coding sequence of pCAG-CBE4max-SpRY-P2A-EGFP was created (CBE-SpRY_U-delp).
The CBE-SpRY-OPT plasmids were created by inserting the 3’UTR+poly-A fragment in the pCAG-CBE4max-SpRY-P2A-EGFP (Addgene #140003) plasmid, and by replacing the CBE4max-SpRY coding sequence with the CBE-SpRY_U-delp fragment. CBE-SpRY-OPT1 and CBE-SpRY-OPT2 plasmids contain a T7 promoter followed by a G and A nt, respectively, allowing alternative capping. Plasmids are available upon request.

sgRNA design

We manually designed sgRNAs targeting the -200 region of the *HBG1/2* promoters and an unrelated genomic site (AAVS1) (**Supplementary Table 1**). To generate the sgRNA expression plasmid, oligonucleotides were annealed to create the sgRNA protospacer and the duplexes were ligated into the Bbs I-digested MA128 plasmid (provided by M. Amendola, Genethon, France). Plasmids are available upon request. For RNA-mediated baseediting we used chemically modified synthetic sgRNAs harboring 2′-O-methyl analogs and 3′-phosphorothioate nonhydrolyzable linkages at the first three 5′ and 3′ nucleotides (Synthego).

mRNA *in vitro* transcription

10 µg of BE-expressing plasmids were digested overnight with 20 Units of a restriction enzyme that cleaves once after the poly-A tail or after the stop codon, for constructs with or without a poly-A tail, respectively. The linearized plasmids were purified using a PCR purification kit (QIAGEN) and were eluted in 30 µl of DNase/RNase-free water. 1 µg of linearized plasmid was used as template for the *in vitro* transcription (ivt) reaction (MEGAscript, Ambion). The ivt protocol was modified as follows. The GTP nucleotide solution was used at a final concentration of 3.0 mM instead of 7.5 mM and the anti-reverse cap analog N7-Methyl-3'-O-Methyl-Guanosine-5'-Triphosphate-5'-Guanosine (ARCA, Trilink) was used at a final concentration of 12.0 mM resulting in a final ratio of Cap:GTP of 4:1 that allows efficient capping of the mRNA. The incubation time for the ivt reaction was reduced to 30 minutes. For constructs without a poly-A tail already included in the plasmid, an additional step of polyadenylation was performed using manufacturer's guidelines (Poly-A tailing kit, Ambion). mRNA was precipitated using lithium chloride and resuspended in TE buffer in a final volume that allowed to achieve a concentration of >1 µg/µl. The mRNA quality was evaluated using Bioanalyzer (Agilent). CBE-SpRY-OPT2 mRNA, containing 5-methoxyuridine, capped with Cap1 analog, and subjected to silica membrane purification, was purchased from Trilink.

Plasmid transfection

K562 and HUDEP-2 cells (10^6 cells/condition) were transfected with 3.6 µg of a base editor-expressing plasmid and 1.2 µg of a sgRNA-containing plasmid. For base editor plasmids that do not express GFP, we co-transfected 250 ng of a GFPmax-expressing plasmid (Lonza). Cells transfected with TE buffer or with a base editor-expressing plasmid only, served as negative controls. We used the AMAXA Cell Line Nucleofector Kit V (VCA-1003) and U-16 and L-29 programs (Nucleofector II) for K562 and HUDEP-2, respectively. 18 h after transfection, transfection efficiency was evaluated by flow cytometry, using the Fortessa X20 (BD Biosciences) or the Gallios (Beckman coulter) flow cytometers. GFP^+ HUDEP-2 cells were sorted 18h after transfection using SH800 Cell Sorter (Sony Biotechnology). The gating strategy used to assess transfection efficiency and to flow sort GFP^+ cells is shown in **Supplementary Figure 18a-d**.

$CD34^+$ HSPCs (10^6 cells/condition) were transfected with 3.6 µg of a base editor-expressing plasmid and 4.5 µg of a sgRNA-containing plasmid or with 4.0 µg of a Cas9-expressing plasmid and 4.9 µg of a sgRNA-containing plasmid. To enrich for edited HSPCs, either we used plasmids that express base editor-GFP fusions or we co-transfected the enzyme-encoding plasmid and 250 ng of a GFPmax-expressing plasmid (Lonza). We used the AMAXA Human CD34 Cell Nucleofector Kit (VPA-1003) and the U-08 program (Nucleofector II). 18 h after transfection, GFP^+ $CD34^+$ HSPCs were sorted based on GFP medium (GFP^+-med) and high (GFP^+-high) expression using SH800 Cell Sorter (Sony Biotechnology). Cells transfected with TE buffer, or with the enzyme-expressing plasmid and plasmid encoding a sgRNA targeting the AAVS1 locus, served as negative controls. The gating strategy used to flow sort GFP^+ cells is shown in **Supplementary Figure 18e-f**.

RNA transfection

10^4 to $2x10^5$ or $2x10^6$ $CD34^+$ HSPCs per condition were transfected with 3.0 µg or 15.0 µg of the enzyme encoding mRNA, respectively, and a synthetic sgRNA at a final concentration of 2.3 µM. We used the P3 Primary Cell 4D-Nucleofector X Kit S or L (Lonza) and the CA137 program (Nucleofector 4D). Untransfected cells or cells transfected with TE buffer or with the enzyme-encoding mRNA only, or with the enzyme-encoding mRNA and a sgRNA targeting the AAVS1 locus, served as negative controls.

Ribonucleoprotein (RNP) transfection

RNP complexes were assembled at room temperature using a 90 µM Cas9-GFP protein and a 180 µM synthetic sgRNA (ratio Cas9:sgRNA of 1:2). $CD34^+$ HSPCs ($2x10^5$ cells/condition) were transfected with RNP complexes using the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza) and the CA137 program (Nucleofector 4D) in the presence of a transfection enhancer (IDT). Untransfected cells or cells transfected with TE buffer or with the enzyme-encoding mRNA only, or with the enzyme-encoding mRNA and a sgRNA targeting the AAVS1 locus, served as negative controls.

HSPC differentiation

Transfected CD34$^+$ HSPCs were differentiated into mature RBCs using a three-phase erythroid differentiation protocol, as previously described[12,65]. During the first phase (day 0 to day 6), cells were cultured in a basal erythroid medium supplemented with 100 ng/ml recombinant human SCF (PeproTech), 5 ng/ml recombinant human IL-3 (PeproTech), 3 IU/ml EPO Eprex (Janssen-Cilag) and 10^{-6} M hydrocortisone (Sigma). During the second phase (day 6 to day 9), cells were co-cultured with MS-5 stromal cells in the basal erythroid medium supplemented with 3 IU/ml EPO Eprex (Janssen-Cilag). During the third phase (day 9 to day 20), cells were co-cultured with stromal MS-5 cells in a basal erythroid medium without cytokines. Heat-inactivated human AB serum was added during the third phase of the differentiation (10%; day 13 to day 20). Erythroid differentiation was monitored by flow cytometry analysis of CD36, CD71, GYPA, BAND3 and α4-Integrin erythroid surface markers and of enucleated cells using the DRAQ5 double-stranded DNA dye. 7AAD was used to identify live cells. The gating strategy used to assess erythroid surface markers and enucleated cells is shown in **Supplementary Figure 19**.

Colony-forming cell (CFC) assay

CD34$^+$ HSPCs were plated at a concentration of $1x10^3$ cells/mL in a methylcellulose-based medium (GFH4435, Stem Cell Technologies) under conditions supporting erythroid and granulocyte/monocyte differentiation. BFU-E and CFU-GM colonies were counted after 14 days. Colonies were randomly picked and collected as bulk populations (containing at least 25 colonies) to evaluate base-editing efficiency, globin expression by RT-qPCR and RP-HPLC and hemoglobin expression by CE-HPLC. BFU-Es were randomly picked and collected as single colonies (around 20 colonies per sample) to evaluate base-editing efficiency and globin expression by RT-qPCR.

Evaluation of editing efficiency

Base-editing efficiency, InDels frequency and the presence of the 4.9-kb deletion were evaluated in K562 and HUDEP-2 cells, 3 days post-transfection, in HSPC-derived erythroid cells at the end of the first phase (day 6) and during the third phase (day 13) of differentiation, in BFU-E and CFU-GM 14 days after plating, in human CD45$^+$ cells sorted from the bone marrow of NBSGW recipient mice, and in spleen and blood derived from the same mice.

Genomic DNA was extracted from control and edited cells using PURE LINK Genomic DNA Mini kit (LifeTechnologies), or Quick-DNA/RNA Miniprep (ZYMO Research), or DNeasy Blood & Tissue Kit (QIAGEN), following manufacturer's instructions. To evaluate base-editing efficiency at sgRNA target sites, we performed PCR followed by Sanger sequencing and EditR analysis (EditR: A Method to Quantify Base Editing from Sanger Sequencing)[66]. TIDE analysis (Tracking of InDels by Decomposition) was also performed in order to evaluate the percentage of insertion and deletion (InDels) in base-edited samples[67]. Supplementary table 2 lists the primers used for PCR.

Digital Droplet PCR (ddPCR) was performed using a primer/probe mix (Biorad) to quantify the frequency of the 4.9-kb deletion. Control primers annealing to *hALB* (located on chr 4) were used as DNA loading control. For the *in vivo* data, the frequency of the 4.9-kb deletion was calculated upon normalization of the ratio of the *HBG1-HBG2* intervening region/*hALB* to the average of the control samples. ddPCR was performed using EvaGreen mix to quantify the frequency of chromosome loss by amplifying a region upstream (*PODL1*) and a region downstream (*CARS*) to the *HBG1/2* promoters, located in the q and p arm of chromosome 11, respectively (adapted from Turchiano et al. 2021[27]). Data was acquired through QX200 analyzer (Bio-Rad) and results were analyzed with QuantaSoftTM Analysis Pro (Bio-Rad). A positive droplet count threshold was set at 30 to allow proper calculation of copy/µL concentration through the application of the Poisson distribution. The frequency of chromosome loss was calculated as the ratio p/q arm (*CARS*/*PODL1*) copy concentrations. Supplementary table 3 lists the primers used for ddPCR.

Genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq)

Human embryonic kidney (HEK) 293T/17 cells (2.5×10^5) were transfected with 500 ng of Cas9-, Cas9-SpG-, or Cas9-SpRY-expressing plasmid, together with 250 ng of each sgRNA–coding plasmid or an empty pUC19 vector (background control), 10 pmol of the bait double-stranded

oligodeoxynucleotide (dsODN) (designed according to the original GUIDE-seq protocol[68]), and 50 ng of a pEGFP-IRES-Puro plasmid, expressing both enhanced GFP (EGFP) and the puromycin resistance genes. One day after transfection, cells were replated and selected with puromycin (1 µg/ml) for 48 hours to enrich for transfected cells. Cells were then collected, and genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) and sheared using the Covaris S200 sonicator to an average length of 500 bp. Library preparation was performed using the original adapters and primers according to previous work[68].

End-repair reaction was performed using NEBNext Ultra End Repair/dA Tailing Module and adaptor ligation using NEBNext® Ultra™ Ligation Module, as previously described[69]. Amplification steps were then performed following the GUIDEseq protocol previously described[68].

Libraries were sequenced with a MiSeq sequencing system (Illumina) using the Illumina MiSeq Reagent kit V2-300 cycles (paired-end sequencing; 2×150-bp). Raw sequencing data (FASTQ files) were analyzed using the GUIDE-seq computational pipeline[70]. Identified sites were considered *bona fide* off-targets if a maximum of seven mismatches against the on-target were present and if they were absent in the background control.

NGS targeted sequencing of on- and off-target sites

On-target sites (*HBG1/2* promoters) were PCR amplified using the Phusion High-Fidelity polymerase (New England BioLabs) and the GC buffer (New England BioLabs). Amplicons were purified using Ampure XP beads (Beckman Coulter). Illumina-compatible barcoded DNA amplicon libraries were prepared using the TruSeq DNA PCR-Free kit (Illumina). PCR amplification was then performed using 1 ng of the double-stranded DNA ligation product and Kapa Taq polymerase reagents (KAPA HiFi HotStart ReadyMix PCR Kit, Kapa Biosystems). After a purification step using Ampure XP beads (Beckman Coulter), libraries were pooled and sequenced using Illumina NovaSeq 6000 system (paired-end sequencing; 2×100-bp). Targeted deep sequencing data were analyzed using CRISPResso2[71].

Off-target sites (identified by GUIDE-seq) were PCR amplified using the Phusion High-Fidelity polymerase (New England BioLabs), the HF buffer (New England BioLabs) and primers containing specific DNA stretches (MR3 for forward primers and MR4 for reverse primers) 5' to the sequence recognizing the off-target site. Amplicons were purified using Ampure XP beads (Beckman Coulter). Illumina-compatible barcoded DNA amplicon libraries were prepared by a second PCR step using the Phusion High-Fidelity polymerase (New England BioLabs), the HF buffer (New England BioLabs) and primers containing Unique Dual Index (UDI) barcodes and annealing to MR3 and MR4 sequences. Libraries were pooled, purified by High Pure PCR Product Purification Kit (Sigma-Aldrich), and sequenced using Illumina NovaSeq 6000 system (paired-end sequencing; 2×150-bp). Targeted deep sequencing data were analyzed using CRISPResso2[71].

Supplementary table 4 lists the primers used for targeted deep sequencing at on- and off-target sites.

RNA-seq

Total RNA was isolated from HD HSPCs 48 h after RNA transfection using the RNeasy Kit (QIAGEN) including a DNAse treatment step. RNA quality was assessed by capillary electrophoresis using High Sensitivity RNA reagents with the Fragment Analyzer (Agilent Technologies) and the RNA concentration was measured by using both Xpose spectrophometry (Trinean) and Fragment Analyzer (Agilent Technologies) capillary electrophoresis.

RNA-seq libraries were prepared starting from 30 ng of total RNA using the Universal Plus mRNA-Seq kit (Nugen) as recommended by the manufacturer. Briefly, mRNA was captured with polyA+ magnetic beads from total RNA. mRNA was chemically fragmented. Single strand and second strand cDNA were produced and then ligated to Illumina compatible adaptors with UDI. To produce oriented RNA-seq libraries, a final step of strand selection was performed. The NuQuant system (Nugen) was used to quantify the RNA-seq libraries. An equimolar pool of the final indexed RNA-Seq libraries was prepared and sequenced using the Illumina NovaSeq 6000 system (paired-end sequencing; 2×100-bp). A total of ~50 millions of passing filter paired-end reads were produced per library.

Read quality was verified using FastQC (v. 0.11.9[72]). Raw reads were trimmed for adapters and low-quality tails (quality < Q20) with BBDuk (v. 38.92[73]); moreover, the first 10 nucleotides were force-trimmed for low quality. Reads shorter than 35 bp after trimming were removed. Reads were subsequently aligned to the human reference genome (hg38) using STAR (v. 2.7.9a[74]). Raw gene counts were obtained in R-4.1.1 using the *featureCounts* function of the *Rsubread* R package (v. 2.6.4[75]) and the GENCODE 38 basic gene annotation for hg38 reference genome. Gene counts were normalized to counts per million mapped reads (CPM) and to fragments per kilobase of exon per million mapped reads (FPKM) using the *edgeR* R package (v. 3.34.1[76]); only genes with a CPM greater than 1 in at least 3 samples were retained for differential analysis. Differential gene expression analysis was performed using the *glmQLFTest* function of the *edgeR* R package, using donor as a blocking variable.

RNA editing analysis was performed accordingly to GATK Best Practices for RNA-seq variant calling (GATK v4.2.2.0). In brief, lane level FASTQ files were two-pass aligned to the hg38 human reference genome with STAR[74] (v2.7.2a) using parameters to specify the ReadGroup and output the aligned BAM file sorted by coordinate. Lane level alignments for each sample were merged and duplicate marked using Picard (v2.25.4). After splitting reads containing Ns in their cigar string because they span splicing sites, base quality recalibration was performed using known variants in dbSNP155. RNA base-editing variant calling was performed using GATK *HaplotypeCaller* only on canonical (1–22, X, Y and M) chromosomes.

Single-nucleotide variants (SNVs) were filtered using the untreated sample as background to identify editing events private to treated samples. Specifically, SNVs without high-confidence reference genotype calls in the untreated experiment were excluded applying the following criteria: coverage ≥ 20 reads, genotype quality ≥ 30, frequency of reference allele ≥ 0.99. Moreover, only SNVs with coverage ≥ 30 reads and genotype quality ≥ 30 were finally retained in the treated samples.

C-to-U editing events comprise C-to-U SNVs called on the positive strand as well as G-to-A SNVs sourced from the negative strand. A-to-I editing events comprise A-to-I SNVs called on the positive strand as well as T-to-C SNVs sourced from the negative strand. SNVs annotation was performed using the Variant Effect Predictor (VEP) tool from Ensembl[77].

Whole exome sequencing

Genomic DNA was extracted from RNA-transfected HD HSPCs using PURE LINK Genomic DNA Mini kit (LifeTechnologies), following manufacturer's instructions. Exome libraries were prepared using the Twist Human RefSeq Exome Kit (36 Mb, Twist Bioscience). Briefly, 100-500 ng of genomic DNA was sheared with an Ultrasonicator (Covaris). A total amount of 50 ng of the fragmented and purified double-stranded DNA was used to prepare the exome libraries as recommended by the manufacturer, but with no initial enzymatic shearing and using adaptors with Unique Molecular Identifier barcodes (IDT). Barcoded exome libraries were pooled and sequenced with the Illumina NovaSeq 6000 system (paired-end sequencing; 2×100-bp). More than 54 million paired-end reads were produced per exome library.

Variant calling was carried out accordingly to GATK Best Practices for germline short variant discovery (GATK v4.2.2.0). In brief, lane level FASTQ files were mapped on the hg38 human genome reference with BWA[78] (v 0.7.17), specifying the ReadGroup. Lane level alignments for each sample were merged, sorted by genomic coordinate and duplicate marked using Picard (v2.25.4). Base quality recalibration was performed specifying the list of target exons (Twist_Exome_RefSeq_targets_hg38.bed) with a padding region of 100 bp. Variant calling was performed using GATK *HaplotypeCaller* only on canonical (1–22, X, Y and M) chromosomes. SNVs were filtered accordingly to the following criteria: coverage ≥ 10 reads and genotype quality ≥ 30. SNVs annotation was performed using the Variant Effect Predictor (VEP) tool from Ensembl[77].

Chromatin Immunoprecipitation assay

GFP^+ high K562 cells were sorted 18 hours post-transfection using the SH800 Cell Sorter (Sony Biotechnology). ChIP experiments were performed in LRF-4C, LRF-8C, LRF-2T, KLF1, Cas9 -197 edited K562 bulk populations and in mock-transfected (transfected with TE buffer) samples. For

each condition, $5x10^7$ cells were cross-linked with 1% formaldehyde for 15 min at room temperature, and the reaction was quenched with glycine at a final concentration of 125 mM. Cross-linked cells were then lysed and sonicated to obtain ~ 200-400 bp fragments of chromatin. Sonicated DNA from each condition was split in two, and each half was pulled down at 4°C overnight using either 25 µg of an antibody against LRF (1 pg/cell of anti-human LRF, 13E9, 14-3309-82, Thermo Fisher-Invitrogen) or 25 µg of an isotype control antibody (1 pg/cell of Armenian hamster IgG isotype control (Arm-IgG), eBio299Arm, 14-4888-85, Thermo Fisher-Invitrogen). Chromatin cross-linking was then reversed, and DNA was eluted at 65°C overnight and purified. Real-time qPCR was performed on ChIP material using the SYBR™ Green PCR Master Mix (Applied Biosystems) and the Viia7 Real-Time PCR system (ThermoFisher Scientific). Supplementary table 5 lists the real-time PCR primers used for ChIP–qPCR.

RT-qPCR

Total RNA was extracted from SCD or β-thalassemic HSPCs differentiated towards the erythroid lineage (day 13) or from healthy donor (HD) HSPCs (12 and 24 hours post-transfection) using RNeasy micro kit (QIAGEN), and from BFU-E pools and single colonies using Quick-DNA/RNA Miniprep (ZYMO Research). RNA was treated with DNase using the DNase I kit (Invitrogen), following manufacturer's instructions. Mature transcripts were reverse-transcribed using SuperScript First-Strand Synthesis System for RT-qPCR (Invitrogen) with oligo (dT) primers. RT-qPCR was performed using the iTaq universal SYBR Green master mix (Biorad) and the Viia7 Real-Time PCR system (ThermoFisher Scientific), or the CFX384 Touch Real-Time PCR Detection System (Biorad). Supplementary table 6 lists the primers used for RT-qPCR.

Flow cytometry analysis

HUDEP-2 were fixed and permeabilized using BD Cytofix/Cytoperm solution (BD Pharmingen) and stained with an antibody recognizing HbF (1/100 APC-conjugated anti-HbF antibody, MHF05, Life Technologies) and an antibody recognizing GYPA erythroid surface marker (1/100 PE-Cy7-conjugated anti-GYPA antibody, 563666, BD Pharmingen). The gating strategy used to assess HbF expression in HUDEP2 cells is shown in **Supplementary Figure 20a-b**.

HSPC-derived erythroid cells were fixed with 0.05 % cold glutaraldehyde and permeabilized with 0.1 % TRITON X-100. After fixation and permeabilization, cells were stained with an antibody recognizing GYPA erythroid surface marker (1/100 PE-Cy7-conjugated anti-GYPA antibody, 563666, BD Pharmingen) and either an antibody recognizing HbF (1/5 FITC-conjugated anti-HbF antibody, clone 2D12 552829 BD), or an antibody recognizing HbS (1/20 anti-HbS antibody, H04181601, BioMedomics) followed by the staining with a secondary antibody recognizing rabbit IgG (1/200 BV421-conjugated anti-rabbit IgG, 565014, BD). The gating strategy used to assess HbF and HbS expression in HSPC-derived erythroid cells is shown in **Supplementary Figure 20c-f**. Flow cytometry analysis of CD36, CD71, GYPA, BAND3 and α4-Integrin erythroid surface markers was performed using a V450-conjugated anti-CD36 antibody (1/20 561535, BD Horizon), a FITC-conjugated anti-CD71 antibody (1/50 555536, BD Pharmingen), a PE-Cy7-conjugated anti-GYPA antibody (1/100 563666, BD Pharmingen), a PE-conjugated anti-BAND3 antibody (1/50 9439, IBGRL) and an APC-conjugated anti-CD49d antibody (1/20 559881, BD). The gating strategy used to assess erythroid surface markers is shown in **Supplementary Figure 19**. Flow cytometry analysis of enucleated or viable cells was performed using double-stranded DNA dyes (DRAQ5, 65-0880-96, Invitrogen and 7AAD, 559925, BD, respectively). The gating strategy used to assess enucleated cells is shown in **Supplementary Figure 19**. Flow cytometry analysis of apoptotic cells was performed using PE-conjugated Annexin V (1/10 559763, BD Pharmingen). The gating strategy used to assess apoptotic cells is shown in **Supplementary Figure 21a-b**. Flow cytometry analysis of reactive oxygen species (ROS) was performed using H_2DCFDA (D399, Invitrogen). The gating strategy used to assess ROS is shown in **Supplementary Figure 21c-d**. Flow cytometry analyses were performed using Fortessa X20 (BD Biosciences) or Gallios (Beckman coulter) flow cytometers. Data were analyzed using the FlowJo (BD Biosciences) software.

RP-HPLC analysis of globin chains

RP-HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). A 250x4.6 mm, 3.6 µm Aeris Widepore column (Phenomenex) was used to separate globin chains by HPLC. Samples were eluted with a gradient mixture of solution A (water/acetonitrile/trifluoroacetic acid, 95:5:0.1) and solution B (water/acetonitrile/trifluoroacetic acid, 5:95:0.1). The absorbance was measured at 220 nm.

CE-HPLC analysis of hemoglobin tetramers
Cation-exchange HPLC analysis was performed using a NexeraX2 SIL-30AC chromatograph and the LC Solution software (Shimadzu). A 2 cation-exchange column (PolyCAT A, PolyLC, Columbia, MD) was used to separate hemoglobin tetramers by HPLC. Samples were eluted with a gradient mixture of solution A (20mM bis Tris, 2mM KCN, pH=6.5) and solution B (20mM bis Tris, 2mM KCN, 250mM NaCl, pH=6.8). The absorbance was measured at 415 nm.

Sickling assay
HSPC-derived mature RBCs obtained at the end of the erythroid differentiation, were incubated under gradual hypoxic conditions (20% O_2 for 20 min; 10% O_2 for 20 min; 5% O_2 for 20 min; 0% O_2 for 60-80 min) and a time course analysis of sickling was performed in real time by video microscopy. Images were captured every 20 min using an AxioObserver Z1 microscope (Zeiss) and a 40x objective. Throughout the time course, images were captured and then processed with ImageJ to determine the percentage of non-sickle RBCs per field of acquisition in the total RBC population. More than 400 cells were counted per condition.

HSPC xenotransplantation in NBSGW mice
NOD.Cg-*Kit*$^{W\text{-}41J}$*Tyr* $^{+}$*Prkdc*^{scid}*Il2rg*^{tm1Wjl}/ThomJ (NBSGW) mice were housed in a pathogen–free facility. Control or edited mobilized healthy donor or non-mobilized SCD CD34^{+} cells (0.4 to 1.2 x 10^6 cells per mouse) were transplanted into nonirradiated NBSGW male and female mice of 5 to 6 weeks of age via retro-orbital sinus injection. NBSGW male and female mice transplanted with non-mobilized SCD CD34^{+} cells were conditioned with busulfan (Sigma, St Louis, MO, US) injected intraperitoneally (10mg/kg body weight/day) 24 h, 48 h and 72 h before transplantation. Neomycin and acid water were added in the water bottle. 16 to 20 weeks after transplantation, NBSGW primary recipients were sacrificed. Cells were harvested from bone marrow, thymus, spleen, and blood, stained with antibodies against murine and human surface markers [murine CD45 (1/50 mCD45-VioBlue), Miltenyi Biotec; human CD45 (1/50 hCD45-APCvio770), Miltenyi Biotec; human CD3 (1/50 CD3-APC), Miltenyi Biotec; human CD14 (1/50 CD14-PECy7), BD Biosciences; human CD15 (1/50 CD15-PE), Miltenyi Biotec; human CD19 (1/100 CD19-BV510); human CD235a (1/50 CD235a-PE), BD Biosciences] and analyzed by flow cytometry using the MACSQuant analyzer (Miltenyi Biotec) and the FlowJo software (BD Biosciences). The gating strategy used to assess chimerism and lineage specific markers is shown in **Supplementary Figure 22**. Human bone marrow CD45^{+} cells were sorted by immunomagnetic selection with AutoMACS (Miltenyi Biotec) after immunostaining with the CD45 MicroBead Kit (Miltenyi Biotec). All experiments and procedures were performed in compliance with the French Ministry of Agriculture's regulations on animal experiments and were approved by the regional Animal Care and Use Committee (APAFIS#2019061312202425_v4). Mice were housed in a temperature (20–22 °C) and humidity (40–50%)-controlled environment with 12 h/12 h light/dark cycle and fed ad libitum with a standard diet.

Statistics and Reproducibility
No statistical method was used to predetermine sample size. We used the minimum number of replicates (n=3) to perform statistics. Biologically independent experiments reported here are from independent (i) splits of each cell type, or (ii) primary cells from different donors, or (iii) mice. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment. Statistical analyses were performed with Prism version 9.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. The RNA-seq data generated in this study and supporting the results of this article have been deposited and are available in the Gene Expression Omnibus repository under the accession number GSE191135 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191135]. The WES data generated in this study and supporting the results of this article have been deposited and are available in the BioProject repository under the accession number PRJNA850889 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA850889]. The GUIDE-seq data generated in this study and supporting the results of this article have been deposited and are available in the BioProject repository under the accession number PRJNA752948 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA752948].

References

1. Kato, G. J. *et al.* Sickle cell disease. *Nat. Rev. Dis. Primer* **4**, 1–22 (2018).
2. Piel, F. B., Steinberg, M. H. & Rees, D. C. Sickle Cell Disease. *N. Engl. J. Med.* **376**, 1561–1573 (2017).
3. Weatherall, D. J. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.* **2**, 245–255 (2001).
4. Cappellini, M. D., Porter, J. B., Viprakasit, V. & Taher, A. T. A paradigm shift on beta-thalassaemia treatment: How will we manage this old disease with new therapies? *Blood Rev.* **32**, 300–311 (2018).
5. Taher, A. T., Weatherall, D. J. & Cappellini, M. D. Thalassaemia. *The Lancet* **391**, 155–167 (2018).
6. Cavazzana, M., Antoniani, C. & Miccio, A. Gene Therapy for β-Hemoglobinopathies. *Mol. Ther. J. Am. Soc. Gene Ther.* **25**, 1142–1154 (2017).
7. Forget, B. G. Molecular basis of hereditary persistence of fetal hemoglobin. *Ann. N. Y. Acad. Sci.* **850**, 38–44 (1998).
8. Wienert, B. *et al.* Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat. Commun.* **6**, 7085 (2015).
9. Wienert, B. *et al.* KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood* **130**, 803–807 (2017).
10. Martyn, G. E. *et al.* A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. *Blood* **133**, 852–856 (2019).
11. Martyn, G. E., Quinlan, K. G. R. & Crossley, M. The regulation of human globin promoters by CCAAT box elements and the recruitment of NF-Y. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1860**, 525–536 (2017).
12. Weber, L. *et al.* Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. *Sci. Adv.* **6**, eaay9392 (2020).
13. Wang, L. *et al.* Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. *Cell Res.* **30**, 276–278 (2020).
14. Li, C. *et al.* In vivo HSPC gene therapy with base editors allows for efficient reactivation of fetal γ-globin in β-YAC mice. *Blood Adv.* **5**, 1122–1135 (2021).
15. Mayuranathan, T. Adenosine Base Editing of γ-Globin Promoters Induces Fetal Hemoglobin and Inhibit Erythroid Sickling. in (ASH, 2020).
16. Traxler, E. A. *et al.* A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat. Med.* **22**, 987–990 (2016).
17. Hoban, M. D. *et al.* Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood* **125**, 2597–2604 (2015).
18. Hoban, M. D. *et al.* CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ cells. *Mol. Ther. J. Am. Soc. Gene Ther.* **24**, 1561–1569 (2016).

19. DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* **8**, 360ra134-360ra134 (2016).
20. Milyavsky, M. *et al.* A Distinctive DNA Damage Response in Human Hematopoietic Stem Cells Reveals an Apoptosis-Independent Role for p53 in Self-Renewal. *Cell Stem Cell* **7**, 186–197 (2010).
21. Cromer, M. K. *et al.* Global Transcriptional Response to CRISPR/Cas9-AAV6-Based Genome Editing in CD34+ Hematopoietic Stem and Progenitor Cells. *Mol. Ther. J. Am. Soc. Gene Ther.* **26**, 2431–2442 (2018).
22. Schiroli, G. *et al.* Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* **24**, 551-565.e8 (2019).
23. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
24. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
25. Boutin, J. *et al.* CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. *Nat. Commun.* **12**, 4922 (2021).
26. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).
27. Turchiano, G. *et al.* Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell* **28**, 1136-1147.e5 (2021).
28. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
29. Antoniou, P., Miccio, A. & Brusson, M. Base and Prime Editing Technologies for Blood Disorders. *Front. Genome Ed.* **3**, 1 (2021).
30. Martyn, G. E. *et al.* Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat. Genet.* **50**, 498–503 (2018).
31. Yang, Y. *et al.* Structural basis for human ZBTB7A action at the fetal globin promoter. *Cell Rep.* **36**, 109759 (2021).
32. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).
33. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
34. Karikó, K., Kuo, A. & Barnathan, E. Overexpression of urokinase receptor in mammalian cells following administration of the in vitro transcribed encoding mRNA. *Gene Ther.* **6**, 1092–1100 (1999).
35. Ross, J. & Sullivan, T. D. Half-lives of beta and gamma globin messenger RNAs and of protein synthetic capacity in cultured human reticulocytes. *Blood* **66**, 1149–1154 (1985).
36. Holtkamp, S. *et al.* Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* **108**, 4009–4017 (2006).
37. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.* **5**, 2108–2116 (1991).
38. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
39. Jiang, T. *et al.* Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. *Nat. Commun.* **11**, 1979 (2020).
40. Vaidyanathan, S. *et al.* Uridine Depletion and Chemical Modification Increase Cas9 mRNA Activity and Reduce Immunogenicity without HPLC Purification. *Mol. Ther. Nucleic Acids* **12**, 530–542 (2018).
41. Yu, W. *et al.* Plexin-B2 Mediates Physiologic and Pathologic Functions of Angiogenin. *Cell* **171**, 849-864.e25 (2017).
42. Liu, Y. *et al.* Latexin Inactivation Enhances Survival and Long-Term Engraftment of Hematopoietic Stem Cells and Expands the Entire Hematopoietic System in Mice. *Stem Cell Rep.* **8**, 991–1004 (2017).

43. Kaur, S. & Roberts, D. D. Divergent modulation of normal and neoplastic stem cells by thrombospondin-1 and CD47 signaling. *Int. J. Biochem. Cell Biol.* **81**, 184–194 (2016).
44. Pinho, S. *et al.* Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. *Dev. Cell* **44**, 634-641.e4 (2018).
45. Li, Y., Turpin, C. P. & Wang, S. Role of thrombospondin 1 in liver diseases. *Hepatol. Res. Off. J. Jpn. Soc. Hepatol.* **47**, 186–193 (2017).
46. Sanghvi, V. R. *et al.* The Oncogenic Action of NRF2 Depends on De-glycation by Fructosamine-3-Kinase. *Cell* **178**, 807-819.e21 (2019).
47. Beeraka, N. M. *et al.* The Taming of Nuclear Factor Erythroid-2-Related Factor-2 (Nrf2) Deglycation by Fructosamine-3-Kinase (FN3K)-Inhibitors-A Novel Strategy to Combat Cancers. *Cancers* **13**, 281 (2021).
48. Magrin, E., Miccio, A. & Cavazzana, M. Lentiviral and genome-editing strategies for the treatment of β-hemoglobinopathies. *Blood* **134**, 1203–1213 (2019).
49. Magrin, E. *et al.* Long-term outcomes of lentiviral gene therapy for the β-hemoglobinopathies: the HGB-205 trial. *Nat. Med.* **28**, 81–88 (2022).
50. Lattanzi, A. *et al.* Development of β-globin gene correction in human hematopoietic stem cells as a potential durable treatment for sickle cell disease. *Sci. Transl. Med.* **13**, eabf2444 (2021).
51. Zeng, J. *et al.* Therapeutic base editing of human hematopoietic stem cells. *Nat. Med.* **26**, 535–541 (2020).
52. Doerfler, P. A. *et al.* Activation of γ-globin gene expression by GATA1 and NF-Y in hereditary persistence of fetal hemoglobin. *Nat. Genet.* **53**, 1177–1186 (2021).
53. Steinberg, M. H., Chui, D. H. K., Dover, G. J., Sebastiani, P. & Alsultan, A. Fetal hemoglobin in sickle cell anemia: a glass half full? *Blood* **123**, 481–485 (2014).
54. Magnani, A. *et al.* Extensive multilineage analysis in patients with mixed chimerism after allogeneic transplantation for sickle cell disease: insight into hematopoiesis and engraftment thresholds for gene therapy. *Haematologica* **105**, 1240–1247 (2020).
55. Smith, A. R. *et al.* Preliminary Results of a Phase 1/2 Clinical Study of Zinc Finger Nuclease-Mediated Editing of BCL11A in Autologous Hematopoietic Stem Cells for Transfusion-Dependent Beta Thalassemia. *Blood* **134**, 3544–3544 (2019).
56. Newby, G. A. *et al.* Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295–302 (2021).
57. Webber, B. R. *et al.* Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat. Commun.* **10**, 5222 (2019).
58. Lattanzi, A. *et al.* Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements. *Mol. Ther.* **27**, 137–150 (2019).
59. Pavani, G. *et al.* Correction of β-thalassemia by CRISPR/Cas9 editing of the α-globin locus in human hematopoietic stem cells. *Blood Adv.* **5**, 1137–1153 (2021).
60. Fu, J. *et al.* Human cell based directed evolution of adenine base editors with improved efficiency. *Nat. Commun.* **12**, 5897 (2021).
61. Jang, H.-K. *et al.* High-purity production and precise editing of DNA base editing ribonucleoproteins. *Sci. Adv.* **7**, eabg2661 (2021).
62. Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
63. Kurita, R. *et al.* Establishment of Immortalized Human Erythroid Progenitor Cell Lines Able to Produce Enucleated Red Blood Cells. *PLoS ONE* **8**, e59890 (2013).
64. Rosello, M. *et al.* Precise base editing for the in vivo study of developmental signaling and human pathologies in zebrafish. *eLife* **10**, e65552 (2021).
65. Giarratana, M.-C. *et al.* Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat. Biotechnol.* **23**, 69–74 (2005).
66. Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.* **1**, 239–250 (2018).
67. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168–e168 (2014).

68. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
69. Nobles, C. L. *et al.* iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. *Genome Biol.* **20**, 14 (2019).
70. Tsai, S. Q., Topkar, V. V., Joung, J. K. & Aryee, M. J. Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.* **34**, 483 (2016).
71. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
72. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
73. BBMap. *SourceForge* https://sourceforge.net/projects/bbmap/.
74. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
75. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**, 923–930 (2014).
76. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).
77. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).

Acknowledgements
This work was supported by state funding from the French National Research Agency (*Agence Nationale de la Recherche*) as part of the *Investissements d'Avenir* program (ANR-10-IAHU-01), the Fondation Bettencourt Schueller, the Paris Ile de France Region (under the "DIM Thérapie génique" initiative), the European Research Council (865797 DITSB), and the AFM-Telethon (grants 22206 and 22399). We thank Christine Bole and the Imagine Institute's genomics facility for generation of the NGS data and Ryo Kurita and Yukio Nakamura for contributing the HUDEP-2 cell line.

Author Contributions
PA designed, conducted experiments, analyzed data and wrote the paper, GH, PM, GF, TF, ACh, LF, JM, MB, GM and MR conducted experiments and analyzed data. JPC, FDB, WEN, MC, VA, JPV, ACe and MA contributed to the design of the experimental strategy. CM, CG and OR analyzed NGS data. AM conceived the study, designed experiments and wrote the paper.

Conflict of interest
PA, MC and AM are named as inventors on a patent describing base-editing approaches for hemoglobinopathies (PCT/EP2021/062633/EP20305488.7: Base-editing approaches for the treatment of beta-hemoglobinopathies). All other authors declare no competing interests.

Figure Legends
Figure 1: LRF BS disruption and KLF1 BS creation in the *HBG1/2* promoters in K562 cells.
a. Schematic representation of the β-globin locus on chromosome 11, depicting the 5' hypersensitive sites of the locus control region (5' LCR HSs; grey boxes), *HBE1*, *HBG2*, *HBG1*, *HBD* and *HBB* genes (colored boxes), the *HBG2* and *HBG1* promoters (white boxes) and the 3' hypersensitive to DNase I site (3'HS). The sequence of the *HBG2* and *HBG1* identical promoters, from –212 to –179 nucleotides upstream of the *HBG* transcription start sites, is shown below. Red and green ovals indicate LRF repressor and KLF1 activator. HPFH mutations identified in the *HBG1* and/or *HBG2* promoters are highlighted by black arrows and HPFH mutations that can be reproduced by ABEs or CBEs are highlighted in green and red, respectively. The percentage of HbF expression in heterozygous HPFH carriers and carriers of SCD (*) or β-thalassemia (**) is indicated in brackets. The sequence of LRF BS upon generation of the LRF 4C, LRF 8C, LRF 2T and KLF1 profiles is presented, and modified bases are highlighted in red and green.

b. ChIP-qPCR analysis of LRF at *HBG1/2* promoters in edited and control (mock transfected) K562 cells. ChIP was performed using an antibody against LRF. *HBG* prom pair of primers was used to amplify the *HBG1/2* promoters. *DEFB122* served as negative control. Data were normalized to the values observed at the *KLF1* locus (positive control). Data are expressed as mean ± SEM (n=3 biologically independent experiments) (left panel). C-G to T-A or A-T to G-C base-editing efficiency of the input and the LRF immunoprecipitated fractions was calculated by the EditR software in samples subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=3 biologically independent experiments) (right panel). * p=0.0140; ** p=0.0040 (Two-way ANOVA with Dunnett correction for multiple comparisons).
Source data are provided as a Source Data file.

Figure 2. **LRF BS disruption and KLF1 BS creation in the *HBG1/2* promoters of SCD HSPC-derived erythroblasts.**
a. Experimental protocol used for base-editing experiments in non-mobilized SCD HSPCs. A BE-, a sgRNA- and a GFP- (optional for enzyme plasmids that do not contain a GFP cassette) expressing plasmid were co-transfected in SCD HSPCs and 18h post-transfection GFP^+ cells were FACS-sorted based on GFP medium (med) and high (high) expression.
b. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software in samples subjected to Sanger sequencing. The LRF 4C editing profile was obtained by pooling data from CBE-NRCH-, CBE-SpG- and CBE-SpRY- treated samples. Data are expressed as mean ± SEM [n=12 (LRF 4C), n=4 (LRF 8C), n=4 (LRF 2T med), n=4 (LRF 2T high), n=3 (KLF1 med), n=4 (KLF1 high) biologically independent experiments, 4 donors].
c. Frequency of InDels, measured by TIDE analysis for control, base- and Cas9- edited samples subjected to Sanger sequencing. The insertion or deletion of a C (±1 nt) in the homopoly-C stretch of the LRF 2T profile was separated from the overall frequency of InDels, as it was considered a sequencing error (**Supplementary Note 2**). Data are expressed as mean ± SEM [n=12 (control), n=12 (LRF 4C), n=4 (LRF 8C), n=4 (LRF 2T med), n=4 (LRF 2T high), n=3 (KLF1 med), n=3 (KLF1 high, n=3 (Cas9 med), n=3 (Cas9 high) biologically independent experiments]. **** $p \leq 0.0001$ (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
d. Frequency of the 4.9-kb deletion, measured by ddPCR, for base- and Cas9- edited samples. Data are expressed as mean ± SEM [n=8 (control), n=12 (LRF 4C), n=5 (LRF 8C), n=3 (LRF 2T med), n=3 (LRF 2T high), n=3 (KLF1 med), n=3 (KLF1 high), n=3 (Cas9 med), n=4 (Cas9 high) biologically independent experiments]. * p=0.0125 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
e. Analysis of HbF and HbS by cation-exchange HPLC in SCD patient RBCs. We calculated the percentage of each Hb type over the total Hb tetramers. The base-editing efficiency is indicated for each sample in the lower part of the panel. Data are expressed as single values or as mean ± SEM [n=4 (control), n=6 (LRF 4C), n=2 (LRF 8C), n=2 (LRF 2T med), n=1 (LRF 2T high), n=1 (KLF1 med), n=1 (KLF1 high), n=1 (Cas9 med), n=2 (Cas9 high) biologically independent experiments, 2 donors]. * p=0.0141 for LRF 4C, or p=0.0380 for LRF 8C; **** $p \leq 0.0001$ (Two-way ANOVA with Dunnett correction for multiple comparisons).
f. Flow cytometry histograms showing the percentage of HbF- and HbS- expressing cells in $GYPA^{+high}$ population for unstained (GYPA stained only), control (transfected with TE buffer for donor 1 and transfected with TE buffer or with CBE-SpRY plasmid and a sgRNA targeting the unrelated *AAVS1* locus for donor 2) and edited samples.
g. Time-course measurement of the frequency of non-sickle cells upon O_2 deprivation in control (transfected with TE buffer for donor 1 and transfected with TE buffer or with CBE-SpRY plasmid and a sgRNA targeting the unrelated *AAVS1* locus for donor 2) and edited samples. Data are expressed as single values or as mean ± SEM [n=2 (control), n=3 (LRF 4C), n=1 (LRF 8C), n=1 (LRF 2T med), n=1 (LRF 2T high), n=1 (KLF1 med), n=1 (KLF1 high), n=1 (Cas9 high) biologically independent experiments, 2 donors].
h. Correlation between *HBG* mRNA relative expression and base-editing efficiency in single BFU-E colonies (1 donor). *HBG* mRNA expression was normalized to *HBA1/2* mRNA and expressed as percentage of the total *HBB*+*HBG* mRNA. Base-editing efficiency was calculated by the

CRISPRESSO 2 software in samples subjected to NGS. Colonies highlighted by a black outline carried InDels. * p≤0.05; ** p≤0.01; *** p≤0.001; **** p≤0.0001 (KLF1: R^2=0.7299, $Y = 0.8896*X + 31.37$, p<0.0001 non-zero slope significancy; LRF 2T: R^2=0.8574, $Y = 0.5977*X + 25.07$, p<0.0001 non-zero slope significancy; LRF 8C: R^2=0.6102, $Y = 0.4824*X + 23.77$, p<0.0001 non-zero slope significancy; LRF 4C: R^2=0.1678, $Y = 0.2529*X + 30.52$, p=0.0522 non-zero slope significancy; Multiple t-test). BFU-Es edited at the *AAVS1* locus were used as negative controls.
i. Frequency of InDels, measured by the CRISPRESSO 2 software, for edited or control (AAVS1) single BFU-E colonies (KLF1 n=17; LRF 2T n=17; LRF 8C n=11; LRF 4C n=15; AAVS1 n=8; 1 donor).
j. Frequency of the 4.9-kb deletion, measured by ddPCR, for edited or control (AAVS1) single BFU-E colonies (KLF1 n=11; LRF 2T n=11; LRF 8C n=9; LRF 4C n=15; AAVS n=8; 1 donor).
k. Frequency of chromosome 11 loss, as indicated by the ratio of *CARS* (p arm) and *PODL1* (q arm), measured by ddPCR, for edited or control (AAVS1) single BFU-E colonies (KLF1 n=16; LRF 2T n=15; LRF 8C n=10; LRF 4C n=14; AAVS1 n=9; 1 donor).
Source data are provided as a Source Data file.

Figure 3. **RNA-mediated base editing in SCD HPSCs**
a. Experimental protocol used for base-editing experiments using BE mRNAs in SCD HSPCs (2 plerixafor-mobilized donors and 1 non-mobilized donor). A BE mRNA and a chemically modified sgRNA were co-transfected in SCD HSPCs. Cells were differentiated into mature RBCs or underwent a CFC assay.
b-e. C-G to T-A (b and c) or A-T to G-C (d-e) base-editing efficiency, calculated by the EditR software in samples subjected to Sanger sequencing at early (Day 6) or late (Day 13) time points during the *in vitro* erythroid differentiation protocol. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).
f. Frequency of InDels, measured by TIDE analysis, for control, base- and Cas9- edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments). **** p≤0.0001 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
g. Frequency of the 4.9-kb deletion, measured by ddPCR, for control, base- and Cas9- edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments). **** p≤0.0001 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
h. CFC frequency for control and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). No statistical differences were observed between control and edited samples (Two-way ANOVA).
i. RT-qPCR analysis of β^S- and γ-globin mRNA levels in SCD patient erythroblasts at day 13 of erythroid differentiation. β^S- and γ-globin mRNA expression was normalized to α-globin mRNA and expressed as percentage of the β^S-+γ- globins mRNA. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). * p=0.0058; *** p=0.0003; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
j. Expression of ^{G}γ-, ^{A}γ-, γ(^{G}γ + ^{A}γ)-globin chains measured by RP-HPLC in SCD patient RBCs. γ-globin expression was normalized to α-globin. Data are expressed as mean ± SEM (n-3 biologically independent experiments, 3 donors). * p=0.0133 for LRF 4C, or p=0.0257 for LRF 8C; *** p=0.0002; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
k. Analysis of HbF and HbS by cation-exchange HPLC in SCD patient RBCs. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). ** p=0.0013; *** p=0.0003; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
l. Flow cytometry histograms showing the percentage of HbF- and HbS-expressing cells in $GYPA^+$ population for unstained (GYPA stained only), control (untreated, or transfected with TE buffer, or transfected with a BE mRNA only, or transfected with a BE mRNA and a sgRNA targeting the unrelated *AAVS1* locus) and edited samples.
m. Frequency of HbF- and HbS- expressing cells in $GYPA^+$ population for unstained, control and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3

donors). * p=0.0105; ** p=0.0046 for LRF 8C and KLF1, or p=0.0011 for LRF 2T; *** p=0.0006 (Two-way ANOVA with Dunnett correction for multiple comparisons).
n. Frequency of sickling cells upon O_2 deprivation in control and edited samples. Data are expressed as single values or as mean ± SEM (n=3 biologically independent experiments, 3 donors).
o. Representative photomicrographs of SCD patient RBCs under hypoxia conditions. Red arrows indicate sickling RBCs and green arrows indicate normal RBCs.
Source data are provided as a Source Data file.

Figure 4. **RNA-mediated base editing in β-thalassemia HSPCs.**
a-b. C-G to T-A (a) or A-T to G-C (b) base-editing efficiency, calculated by the EditR software in β-thalassemic erythroblasts subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=2 biologically independent experiments, peripheral blood-derived non-mobilized HSPCs from 2 donors harboring the CD39/IVS1-110 G>A and CD 8/9 +G/IVS1-110 G>A mutations, respectively).
c. Frequency of InDels, measured by TIDE analysis, for base-edited samples subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=2 biologically independent experiments).
d. CFC frequency for control (transfected with TE buffer) and base-edited samples. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). No statistical differences were observed between control and edited samples (Two-way ANOVA).
e. RT-qPCR analysis of β-like globin mRNA levels in β-thalassemia patient erythroblasts at day 13 of erythroid differentiation. β-like globin mRNA expression was normalized to α-globin mRNA. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). * p=0.0270; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
f. Analysis of HbF, HbA and HbA_2 by cation-exchange HPLC in β-thalassemia patient RBCs. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
g. Frequency of HbF-expressing cells in the $GYPA^+$ population for control and edited samples, as measured by flow cytometry. Data are expressed as mean ± SEM (n=4 biologically independent experiments, 2 donors). ** p=0.0098 for LRF 8C, or p=0.0033 for KLF1 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons). Representative flow cytometry histograms showing HbF^+ cells in $GYPA^+$ populations for control and base-edited samples are presented below the graph.
h. Expression of β-, δ-, ^{G}γ-, ^{A}γ- and γ- (^{G}γ- + ^{A}γ-) globin chains measured by RP-HPLC in β-thalassemia patient RBCs. β-like globin expression was normalized to α-globin. The ratio α/non-α globins is reported on top of the graph. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). * p=0.0148 (Two-way ANOVA with Dunnett correction for multiple comparisons).
i. Analysis of α-globin precipitates by cation-exchange HPLC in β-thalassemia patient RBCs. We calculated the proportion of α-globin precipitates over the total Hb tetramers. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). *** p=0.0009 for LRF 8C, or p=0.0004 for KLF1 (One-way ANOVA with Dunnett correction for multiple comparisons).
j. Frequency of enucleated cells at day 6, 9, 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of DRAQ5 nuclear staining in control and edited samples. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). Representative flow cytometry histograms showing the $DRAQ5^-$ cell population for unstained, control and edited samples are presented below the graph. *** p=0.0003; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
k. Cell size of enucleated cells ($DRAQ5^-$) at day 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry using the median of forward scatter (FSC) intensity, and normalized to HD RBCs. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). Representative flow cytometry contour plots showing the FSC of $DRAQ5^-$ cell population for control and edited samples are reported below the graph. ** p=0.0033 for D13, or p=0.0021 for D16, or p=0.0034 for D20; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).

l-n. Frequency of $CD36^+$ (l), $CD71^+$ (m) and $GYPA^+$ (n) cells at day 6, 9, 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of CD36, CD71 and GYPA erythroid markers. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). Representative flow cytometry histograms showing the $CD36^+$ (l), $CD71^+$ (m) and $GYPA^+$ (n) cell populations for unstained, control and edited samples are presented below the graph. * p=0.0241 for CD36/D16, or p=0.0190 for CD36/D20, or p=0.0307 for CD71; ** p=0.0072 for CD36, or p=0.0020 for CD71/LRF 8C, or p=0.0028 for CD71/KLF1; *** p=0.0002; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
o. Frequency of α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ in $7AAD^-$/$GYPA^+$ cells at day 6, 9, 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of α4-Integrin and BAND3 erythroid markers. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). Representative flow cytometry contour plots showing the α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ cell populations for unstained, control and edited samples are reported below the graph.
p. Frequency of apoptotic cells ($AnnexinV^+$-cells) in control and edited samples at day 13 of erythroid differentiation, as measured by flow cytometry. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). Representative flow cytometry contour plots showing the $AnnexinV^+$ cell populations for unstained, control and edited samples are reported on the right side of the graph. ** p=0.0094 for LRF 8C, or p=0.0014 for KLF1 (One-way ANOVA with Dunnett correction for multiple comparisons)
q. Frequency of ROS-containing cells ($DCFDA^+$ cells) in control and edited samples at day 20 of erythroid differentiation, as measured by flow cytometry analysis in $DRAQ^+$ and $DRAQ^-$ cells. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). Representative flow cytometry contour plots showing the $DCFDA^+$ cell populations for unstained, control and edited enucleated samples are reported on the right side of the graph.
Source data are provided as a Source Data file.

Figure 5. **DNA damage and immune responses in HSPCs upon RNA-mediated base editing and RNP-mediated Cas9 treatment.**
a. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software in mobilized HD HSPC samples subjected to Sanger sequencing 6 days post-transfection. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).
b. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software in mobilized HD HSPC samples subjected to Sanger sequencing 12h-, 24h-, 48h-, and 6 days post-transfection. Data are expressed as mean ± SEM (n=2-3 biologically independent experiments, 2-3 donors).
c. Frequency of InDels, measured by TIDE analysis, for control, base- and Cas9-edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). *** p=0.0005 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
d. Frequency of the 4.9-kb deletion, measured by ddPCR, for control, base- and Cas9- edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). **** p≤0.0001 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
e. RT-qPCR analysis of genes activated by RNA stimuli, 24 hours post-transfection in HD HSPCs. *TNF-α*, *IL-6*, *IL-12*, *IFN-α*, *IFN-β*, *TLR7*, *TLR8*, *RIG-I* mRNA expression was normalized to GAPDH mRNA. LPS-activated macrophages were used as positive control. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). **** p≤0.0001 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
f. RT-qPCR analysis of *CDKN1* (p21), 24, 48 and 72 hours post-transfection in HD HSPCs. *CDKN1* mRNA expression was normalized to GAPDH mRNA. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). * p=0.0292 for 24h, or p=0.0259 for 48h, or p=0.0131 for 72h; ** p=0.0019; **** p≤0.0001 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
Source data are provided as a Source Data file.

Figure 6. Transcriptomic analysis of HD HSPCs after RNA-mediated base editing and RNP-mediated Cas9 treatment.
a. RNA-seq analysis was performed 48h after transfection. Mean-difference plots show differentially expressed genes of edited samples over control samples. Genes that are not statistically significant (false discovery rate [FDR] > 0.05) differentially expressed are depicted by black dots. Genes that are statistically significant (FDR < 0.05) upregulated or downregulated are depicted by red and blue dots, respectively. The enzyme used and the profile generated are indicated on top of each plot.
b. Expression of statistically significant (FDR < 0.05) upregulated or downregulated genes, as measured by the log2-fold change of FPKM of edited samples over control samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).
c. Strip plots showing the variant allele frequency of A>G mutations or C>T mutations in RNA observed in HSPCs obtained from three different HD. The total number of mutations are indicated above each sample.
Source data are provided as a Source Data file.

Figure 7. sgRNA-dependent DNA off-target activity of base-editing system.
a-c. sgRNA-dependent off-target DNA sites, as evaluated by GUIDE-seq analysis, of LRF_bs_3 (LRF 4C) (**a**), LRF_bs_2 (LRF 8C) (**b**), and KLF1_bs_1 (LRF 2T and KLF1) (**c**) sgRNAs in HEK293T cells. sgRNAs were coupled with a Cas9 nuclease corresponding to the Cas9 nickase included in the base editor (Cas9-SpRY for LRF_bs_3 and LRF_bs_2 sgRNAs, and Cas9 for KLF1_bs_1 sgRNA). The protospacer targeted by each sgRNA and the PAM are reported on top of each panel, followed by the off-target sites and their mismatches with the on-target (highlighted in color). The number of sequencing reads, the chromosomal coordinates (Human GRCh37/hg19) and the site of each off-target are reported.
d-f. Frequency of C-G to T-A (**d** and **e**) or A-T to G-C (**f**) base-editing conversion at on-target and off-target (OT) sites, for control and LRF 4C (**d**), LRF 8C (**e**), LRF 2T (**f**), and KLF1 (**f**) samples, as measured by targeted NGS sequencing. Data are expressed as individual values and median (n=3 biologically independent experiments, 3 donors). * p=0.0107 for **d**, or p=0.0299 for **f**; ** p=0.0022; **** $p \leq 0.0001$ [Two-way ANOVA with Sidak (d and f) or Tukey (f) correction for multiple comparisons].
g. Venn diagrams showing the overlapping of C>T or A>G single nucleotide variants in exons, in control (Ctr), CBE-SpRY-OPT2-, or ABEmax- treated HSPCs obtained from three different HD.
Source data are provided as a Source Data file.

Figure 8. RNA-mediated base editing of the -200 region of *HBG* promoters in repopulating HSCs.
a. Experimental protocol of HSPC xenotransplantation in NBSGW mice. G-CSF-mobilized HD HSPCs or non-mobilized SCD HSPCs were subjected to RNA-mediated base editing. A BE mRNA and a chemically modified sgRNA were co-transfected in HSPCs and cells were xenotransplanted into NBSGW immunodeficient mice.
b. Engraftment of human cells in NBSGW mice transplanted with control (mock-transfected, transfected with CBE or ABE mRNA alone) and edited (LRF 8C or KLF1) mobilized HD or SCD HSPCs [HD: n=4 (mock, CBE mRNA, LRF 8C, ABE mRNA), n=3 (KLF1) mice per group; SCD: n=3 (mock), n=2 (LRF 8C), n=5 (KLF1) mice per group] 16 to 20 weeks post-transplantation. Engraftment is represented as percentage of human $CD45^+$ cells in the total murine and human $CD45^+$ cell population, in bone marrow (BM), spleen, thymus, and peripheral blood. Each data point represents an individual mouse. Data are expressed as mean ± SEM.
c. Human hematopoietic progenitor content in BM human $CD45^+$ cells derived from mice transplanted with control and edited HSPCs [HD: n=4 (mock), n=3 (CBE), n=4 (LRF 8C), n=3 (ABE), n=2 (KLF1) mice per group; SCD: n=5 (mock), n=1 (LRF 8C), n=2 (KLF1) mice per group]. We plotted the percentage of human $CD45^+$ cells giving rise to BFU-E and CFU-GM. Data are expressed as mean ± SEM.
d. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software, in input, bone marrow-, spleen-, BFU-E-, CFU-GM-, and peripheral blood-derived HD and SCD human samples

subjected to Sanger sequencing. Data are expressed as mean ± SEM [HD-LRF 8C: n=3 (Input) biologically independent experiments, n=4 (Bone Marrow and Blood), n=3 (Spleen, BFU-E, CFU-GM) mice per group; HD-KLF1: n=3 (Input) biologically independent experiments, n=3 (Bone Marrow and Spleen), n=2 (BFU-E, CFU-GM and Blood) mice per group; SCD-LRF 8C: n=2 (Input) biologically independent experiments, n=2 (Bone Marrow, Spleen and Blood), n=1 (BFU-E and CFU-GM) mice per group; SCD-KLF1: n=2 (Input) biologically independent experiments, n=6 (Bone Marrow and Spleen), n=2 (BFU-E and CFU-GM), n=5 (Blood) mice per group]. The frequency of base editing in input cells was calculated in cells cultured in the HSPC medium (pointing-up triangle), in liquid erythroid cultures (rhombus), BFU-E (square) and CFU-GM (pointing-down triangle) colonies. Each data point (circle) represents an individual mouse. HD/LRF8C: ** p=0.021 for Bone marrow, or p=0.0094 for Spleen, or p=0.0037 for BFU-E, or p=0.0059 for CFU-GM, or p=0.0042 for Blood. HD/KLF1: * p=0.0136; ** p=0.0065; *** p=0.0009. SCD/LRF8C: ** p=0.0056 for Bone marrow, or p=0.012 for BFU-E; *** p=0.0004 for Spleen, or p=0.0010 for CFU-GM, or p=0.0001 for Blood. SCD/KLF1: ** p=0.0040; *** p=0.0009; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).

e. Base-editing profile for LRF 8C and KLF1 samples, calculated using EditR software, in input, bone marrow-, spleen- and peripheral blood-derived human samples subjected to Sanger sequencing. Data are expressed as mean ± SEM [LRF 8C: n=5 (Input) biologically independent experiments, n=6 (Bone Marrow and Blood), n=5 (Spleen and CFU-GM), n=4 (BFU-E) mice per group; KLF1: n=5 (Input) biologically independent experiments, n=9 (Bone Marrow and Spleen), n=8 (Blood) n=4 (BFU-E and CFU-GM) mice per group].

f. Frequency of InDels, measured by TIDE analysis, in input, bone marrow-, spleen- and peripheral blood-derived human samples subjected to Sanger sequencing. Data are expressed as mean ± SEM [Input: n=7 (Control), n=5 (LRF 8C and KLF1) biologically independent experiments; Bone Marrow: n=11 (Control), n=6 (LRF 8C), n=9 (KLF1) mice per group; Spleen: n=11 (Control), n=5 (LRF 8C), n=9 (KLF1) mice per group; Blood: n=11 (Control), n=4 (LRF 8C), n=8 (KLF1) mice per group; BFU-E: n=9 (Control), n=4 (LRF 8C), n=4 (KLF1) mice per group; CFU-GM: n=10 (Control), n=5 (LRF 8C), n=4 (KLF1) mice per group].

g. Frequency of the 4.9-kb deletion, measured by ddPCR, in input samples (left panel). Frequency of mice that bear the 4.9-kb deletion in bone marrow-derived human $CD45^+$ cells (right panel). Data are expressed as mean ± SEM (n=11 (control), n=5 (LRF 8C and KLF1) biologically independent experiments for left panel and n=6-9 mice per group for right panel). *** p=0.0002 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

h. A-T to G-C base-editing efficiency at on- and off-target sites, calculated by the EditR software, in input, bone marrow- and spleen-derived HD and SCD human samples subjected to Sanger sequencing. Data are expressed as mean ± SEM [ctr: n=3 biologically independent experiments; Input: n=3 (HD), n=2 (SCD) biologically independent experiments; BM: n=3 (HD), n=6 (ON), n=5 (OT4) mice per group]. ns p=0.5920 for HD, or p>0.9999 for SCD; * p=0.0104; ** p=0.0033 (Two-way ANOVA with Dunnett correction for multiple comparisons).

i. Base editing in single BFU-E colonies derived from engrafting HD HSPCs, calculated by the EditR software. We plotted the frequency of BFU-E with 0, 1, 2, 3 or 4 edited *HBG* promoters, the frequency of BFU-E with 0, 1 or 2 edited OT4 alleles and the frequency of BFU-E edited only at *HBG* promoters or OT4 or at both *HBG* promoters and OT4 (n=42 BFU-E obtained from 2 different mice).

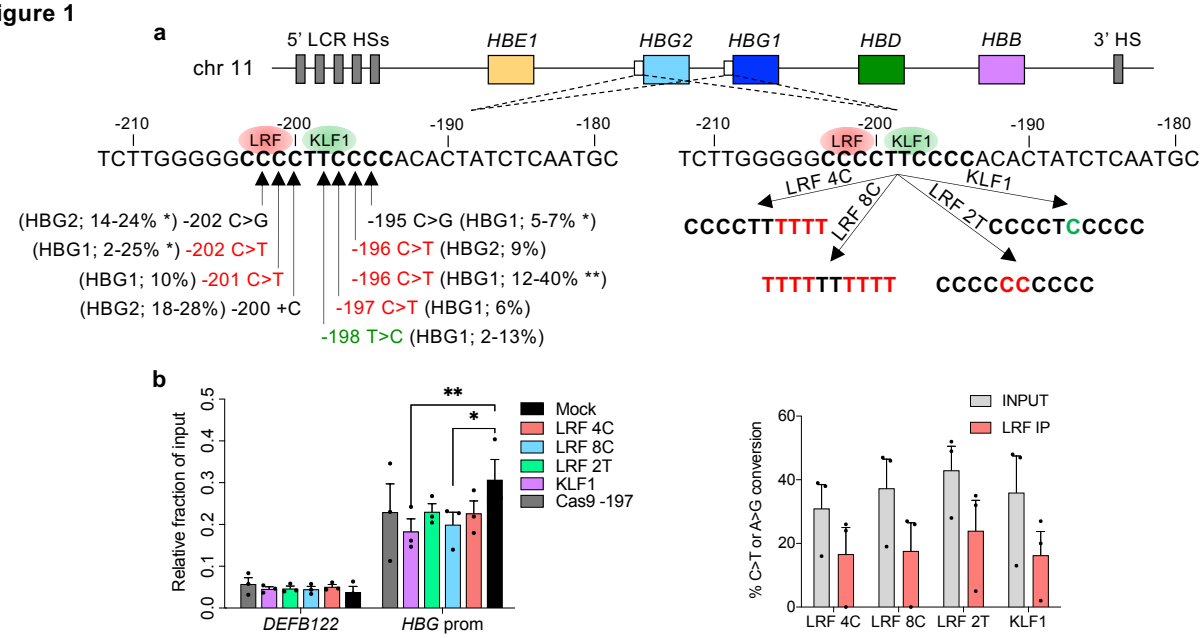
j. Frequency of chromosome 11 loss, as indicated by the ratio of *CARS* (p arm) and *PODL1* (q arm), measured by ddPCR, for edited or control (AAVS1) single BFU-E colonies (KLF1 n=29 biologically independent colonies; AAVS1 n=6 biologically independent colonies; 1 donor).

k. RT-qPCR analysis of β-like globin mRNA levels in bone marrow-derived BFU-E. β-like globins mRNA expression was normalized to α-globin mRNA. Data are expressed as mean ± SEM [n=7 (Mock), n=3 (ABE mRNA), n=4 (KLF1) biologically independent experiments; HD (black circles) and SCD (empty circles) samples]. **** p≤0.0001 (Two-way ANOVA with Sidak correction for multiple comparisons).

l. Correlation between γ-globin mRNA relative expression and base-editing efficiency in bone marrow-derived single BFU-E (n=69). γ-globin mRNA expression was normalized to α-globin mRNA

and expressed as percentage of the total β- and γ- globin mRNA. Base-editing efficiency was calculated by the EditR software in samples subjected to Sanger sequencing. (R^2=0.4263, Y = 0.5328*X + 51.81, p<0.0001 non-zero slope significancy; Simple linear regression).
Source data are provided as a Source Data file.

Figure 1


a
chr 11
5' LCR HSs
HBE1
HBG2
HBG1
HBD
HBB
3' HS
-210
-200
-190
-180
LRF
KLF1
TCTTGGGGGCCCCTTCCCCACACTATCTCAATGC
(HBG2; 14-24% *) -202 C>G
(HBG1; 2-25% *) -202 C>T
(HBG1; 10%) -201 C>T
(HBG2; 18-28%) -200 +C
-195 C>G (HBG1; 5-7% *)
-196 C>T (HBG2; 9%)
-196 C>T (HBG1; 12-40% **)
-197 C>T (HBG1; 6%)
-198 T>C (HBG1; 2-13%)
LRF 4C
LRF 8C
LRF 2T
KLF1
CCCCTTTTTT
TTTTTTTTTT
CCCCCCCCCC

b
Relative fraction of input
0.0
0.1
0.2
0.3
0.4
0.5
**
*
Mock
LRF 4C
LRF 8C
LRF 2T
KLF1
Cas9 -197
DEFB122
HBG prom

% C>T or A>G conversion
0
20
40
60
INPUT
LRF IP
LRF 4C
LRF 8C
LRF 2T
KLF1

Figure 2

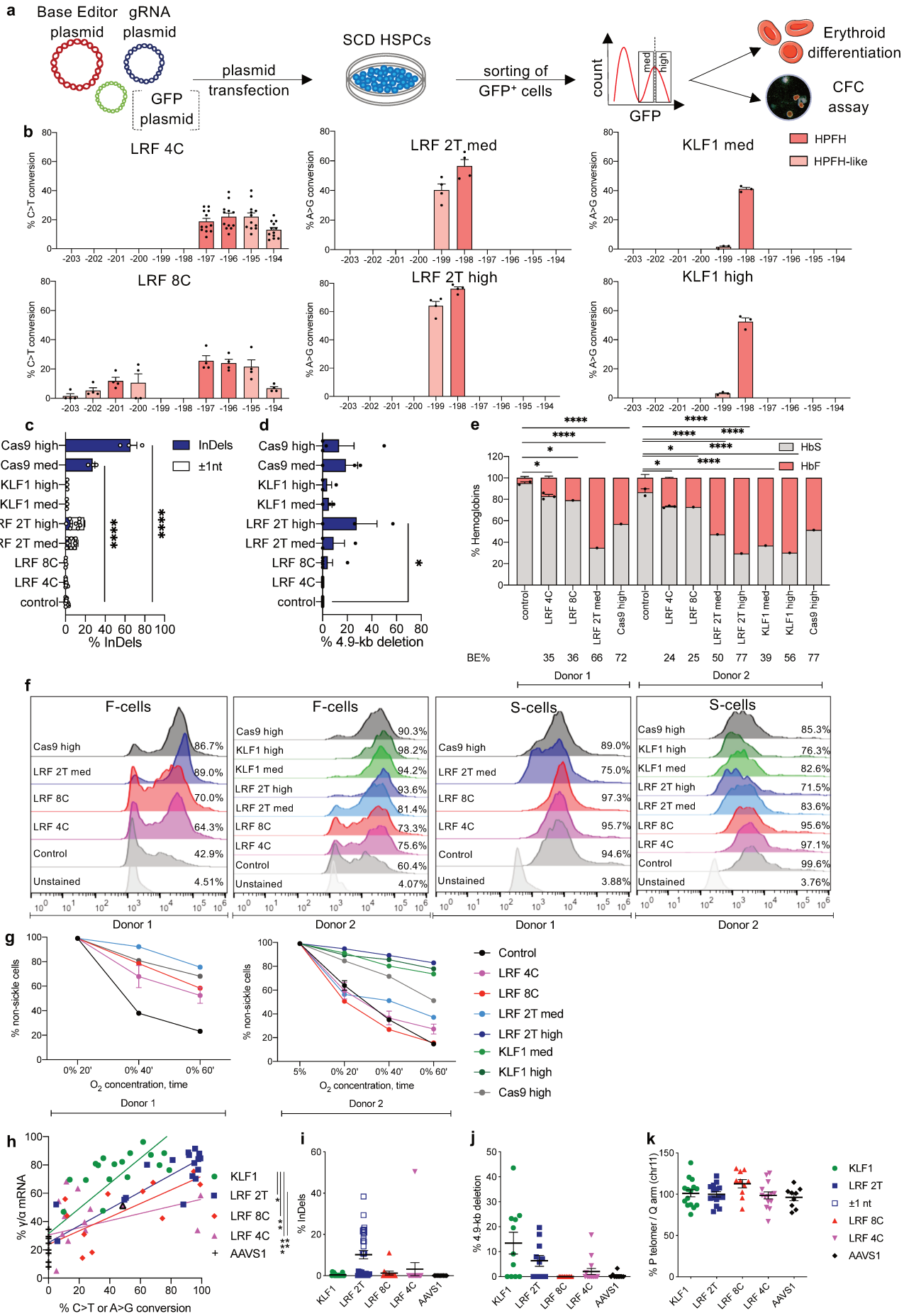

a
Base Editor plasmid
gRNA plasmid
GFP plasmid
plasmid transfection
SCD HSPCs
sorting of GFP+ cells
count
med
high
GFP
Erythroid differentiation
CFC assay
b
LRF 4C
LRF 2T med
KLF1 med
HPFH
HPFH-like
LRF 8C
LRF 2T high
KLF1 high
% C>T conversion
% A>G conversion
c
Cas9 high
Cas9 med
KLF1 high
KLF1 med
LRF 2T high
LRF 2T med
LRF 8C
LRF 4C
control
InDels
±1nt
% InDels
d
% 4.9-kb deletion
e
% Hemoglobins
HbS
HbF
BE%
35 36 66 72 24 25 50 77 39 56 77
f
F-cells
S-cells
Donor 1
Donor 2
Unstained
g
% non-sickle cells
O2 concentration, time
5% 0% 20' 0% 40' 0% 60'
h
% γ/α mRNA
% C>T or A>G conversion
AAVS1
i
% InDels
j
% 4.9-kb deletion
k
% P telomer / Q arm (chr11)

Figure 3

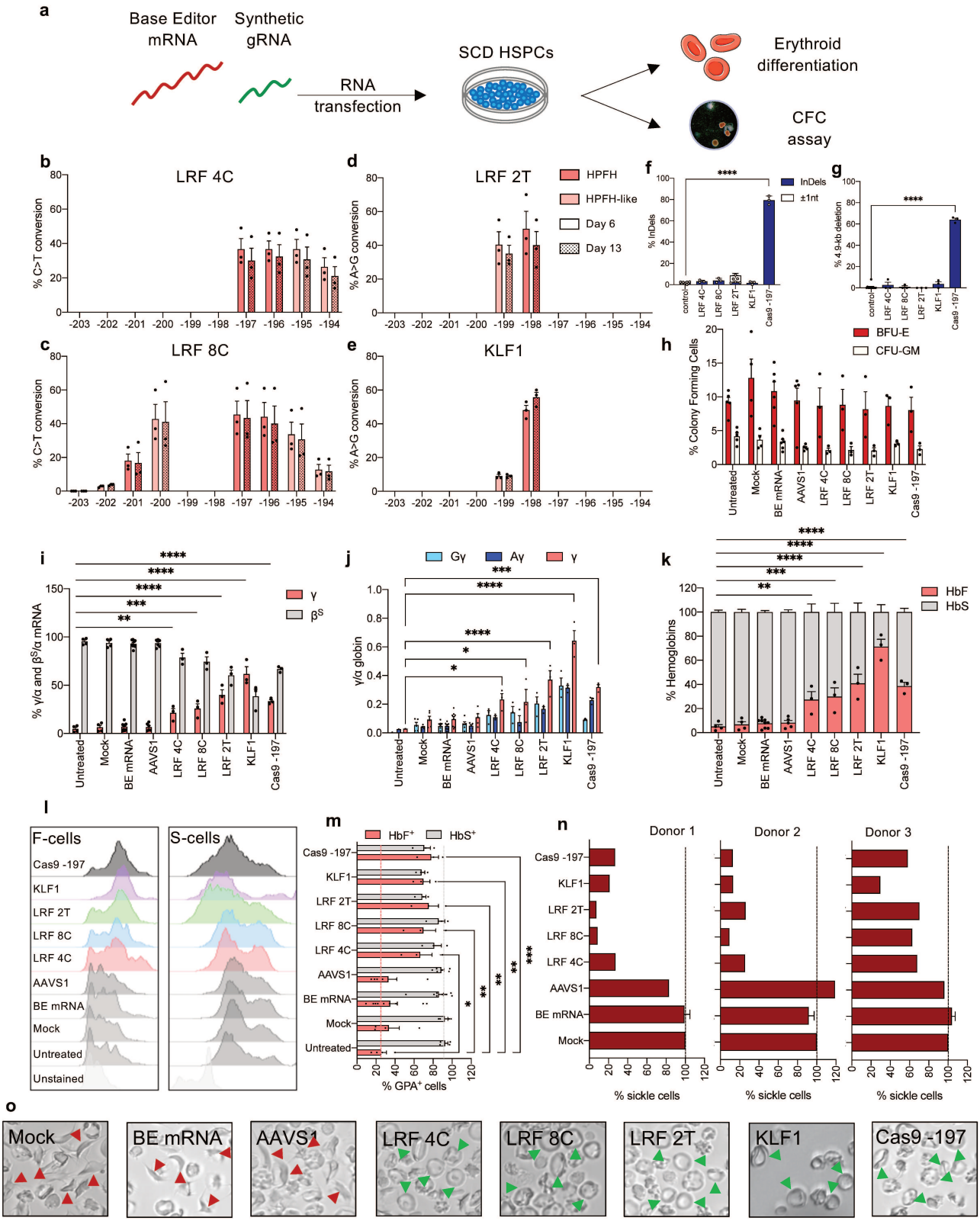

a
Base Editor mRNA
Synthetic gRNA
RNA transfection
SCD HSPCs
Erythroid differentiation
CFC assay
b
LRF 4C
% C>T conversion
d
LRF 2T
% A>G conversion
HPFH
HPFH-like
Day 6
Day 13
f
% InDels
InDels
±1nt
g
% 4.9-kb deletion
c
LRF 8C
e
KLF1
h
% Colony Forming Cells
BFU-E
CFU-GM
Untreated
Mock
BE mRNA
AAVS1
LRF 4C
LRF 8C
LRF 2T
KLF1
Cas9 -197
i
% γ/α and βS/α mRNA
γ
βS
j
Gγ
Aγ
γ
γ/α globin
k
% Hemoglobins
HbF
HbS
l
F-cells
S-cells
Unstained
m
HbF+
HbS+
% GPA+ cells
n
Donor 1
Donor 2
Donor 3
% sickle cells
o
Mock
BE mRNA
AAVS1
LRF 4C
LRF 8C
LRF 2T
KLF1
Cas9 -197

Figure 4

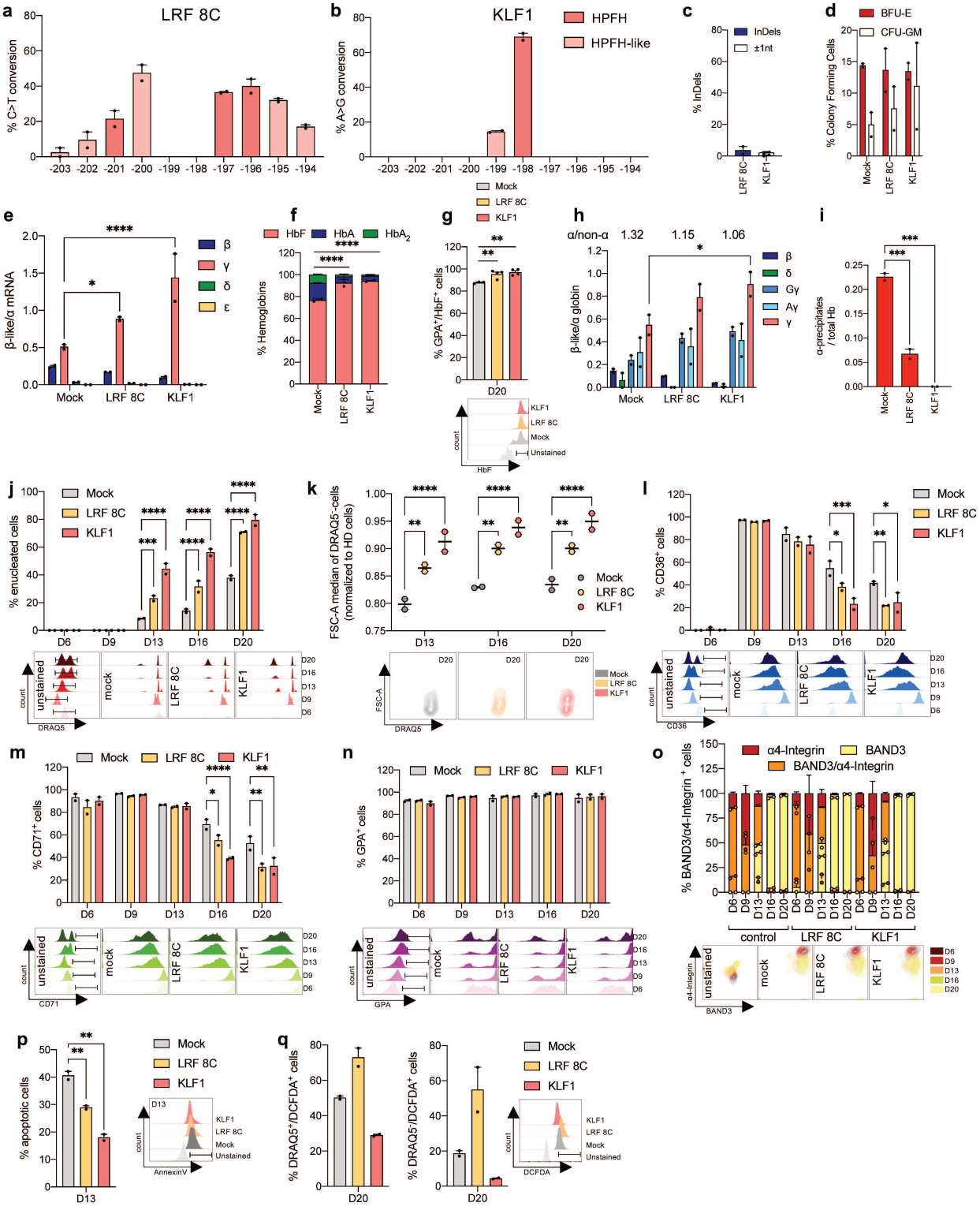

a
LRF 8C
% C>T conversion
-203 -202 -201 -200 -199 -198 -197 -196 -195 -194
b
KLF1
HPFH
HPFH-like
% A>G conversion
Mock
LRF 8C
KLF1
c
InDels
±1nt
% InDels
d
BFU-E
CFU-GM
% Colony Forming Cells
e
β-like/α mRNA
β
γ
δ
ε
f
HbF
HbA
HbA2
% Hemoglobins
g
% GPA+/HbF+ cells
D20
count
Unstained
h
α/non-α 1.32 1.15 1.06
β-like/α globin
β
δ
Gγ
Aγ
γ
i
α-precipitates / total Hb
j
% enucleated cells
D6 D9 D13 D16 D20
unstained
mock
DRAQ5
k
FSC-A median of DRAQ5+cells (normalized to HD cells)
FSC-A
l
% CD36+ cells
CD36
m
% CD71+ cells
CD71
n
% GPA+ cells
GPA
o
α4-Integrin
BAND3
BAND3/α4-Integrin
% BAND3/α4-Integrin+ cells
control
LRF 8C
KLF1
p
% apoptotic cells
AnnexinV
q
% DRAQ5+/DCFDA+ cells
DCFDA

Figure 5

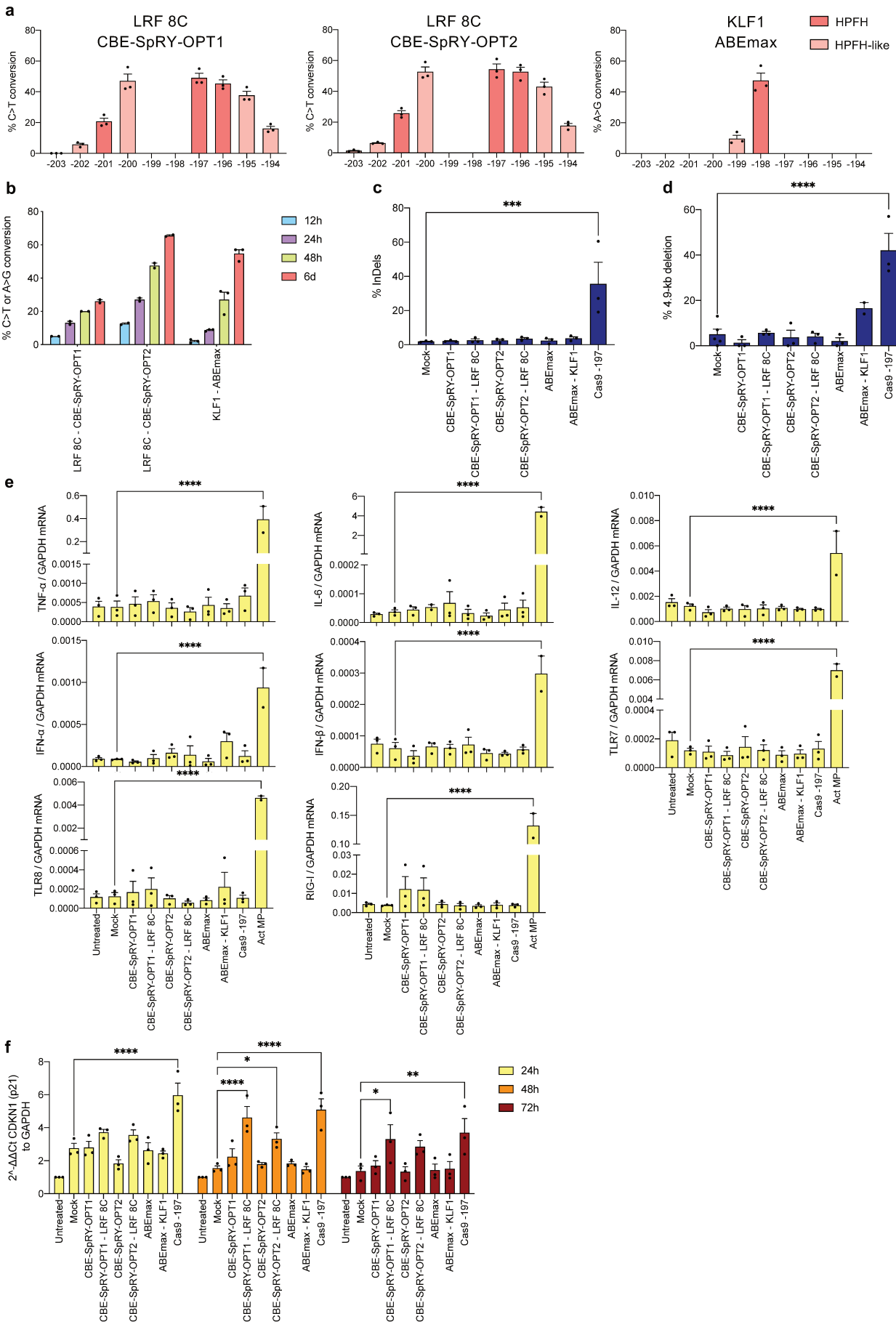

a
LRF 8C
CBE-SpRY-OPT1
% C>T conversion
LRF 8C
CBE-SpRY-OPT2
% C>T conversion
KLF1
ABEmax
% A>G conversion
HPFH
HPFH-like
-203 -202 -201 -200 -199 -198 -197 -196 -195 -194
b
% C>T or A>G conversion
12h
24h
48h
6d
LRF 8C - CBE-SpRY-OPT1
LRF 8C - CBE-SpRY-OPT2
KLF1 - ABEmax
c
% InDels
***
Mock
CBE-SpRY-OPT1
CBE-SpRY-OPT1 - LRF 8C
CBE-SpRY-OPT2
CBE-SpRY-OPT2 - LRF 8C
ABEmax
ABEmax - KLF1
Cas9 -197
d
% 4.9-kb deletion
****
e
TNF-α / GAPDH mRNA
IL-6 / GAPDH mRNA
IL-12 / GAPDH mRNA
IFN-α / GAPDH mRNA
IFN-β / GAPDH mRNA
TLR7 / GAPDH mRNA
TLR8 / GAPDH mRNA
RIG-I / GAPDH mRNA
Untreated
Mock
CBE-SpRY-OPT1
CBE-SpRY-OPT1 - LRF 8C
CBE-SpRY-OPT2
CBE-SpRY-OPT2 - LRF 8C
ABEmax
ABEmax - KLF1
Cas9 -197
Act MP
f
2^-ΔΔCt CDKN1 (p21) to GAPDH
24h
48h
72h
*
**

Figure 6

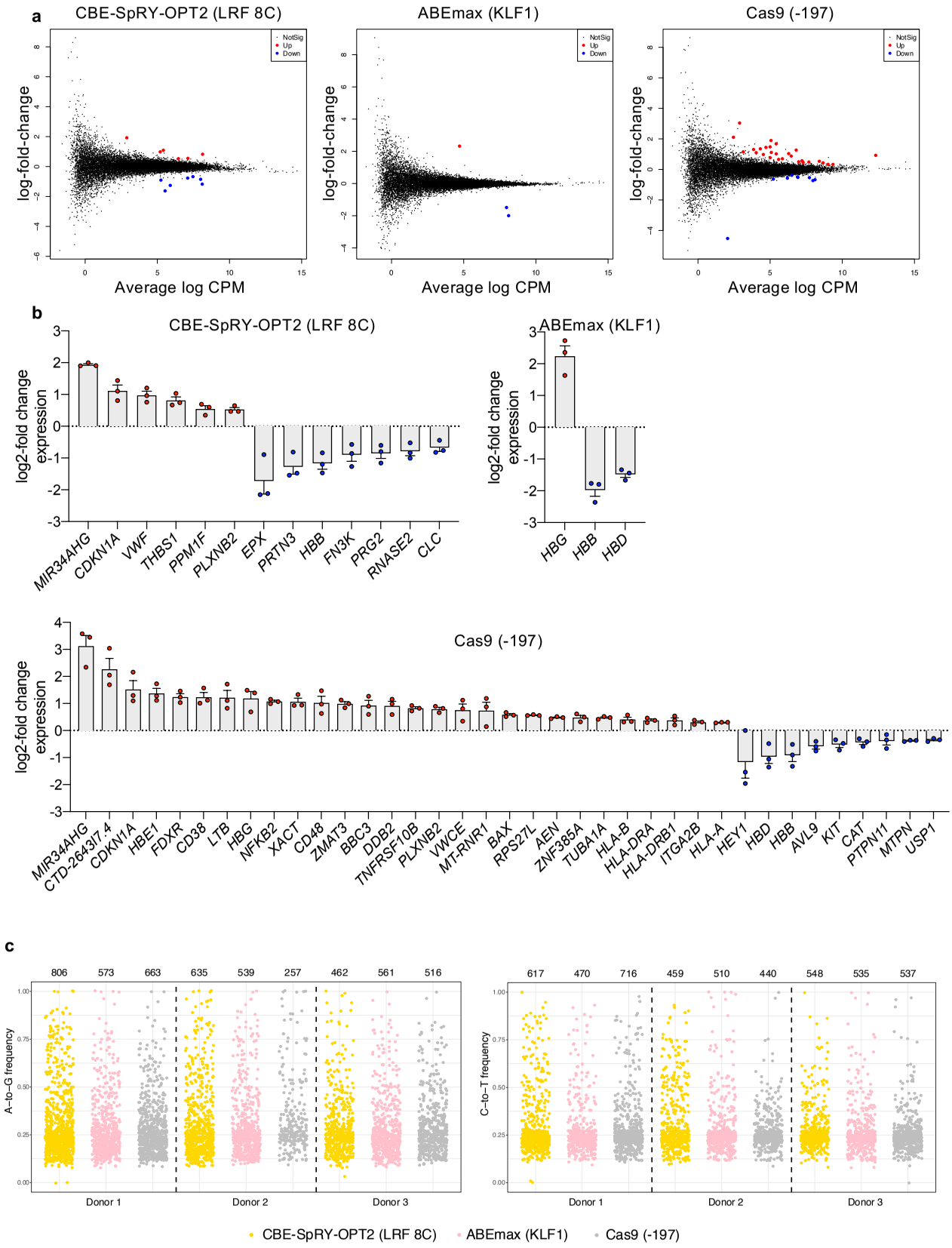

a
CBE-SpRY-OPT2 (LRF 8C)
ABEmax (KLF1)
Cas9 (-197)
NotSig
Up
Down
log-fold-change
Average log CPM
b
CBE-SpRY-OPT2 (LRF 8C)
ABEmax (KLF1)
log2-fold change expression
MIR34AHG
CDKN1A
VWF
THBS1
PPM1F
PLXNB2
EPX
PRTN3
HBB
FN3K
PRG2
RNASE2
CLC
HBG
HBB
HBD
Cas9 (-197)
MIR34AHG
CTD-2643I7.4
CDKN1A
HBE1
FDXR
CD38
LTB
HBG
NFKB2
XACT
CD48
ZMAT3
BBC3
DDB2
TNFRSF10B
PLXNB2
VWCE
MT-RNR1
BAX
RPS27L
AEN
ZNF385A
TUBA1A
HLA-B
HLA-DRA
HLA-DRB1
ITGA2B
HLA-A
HEY1
HBD
HBB
AVL9
KIT
CAT
PTPN11
MTPN
USP1
c
806 573 663 635 539 257 462 561 516
617 470 716 459 510 440 548 535 537
A–to–G frequency
C–to–T frequency
Donor 1
Donor 2
Donor 3
CBE-SpRY-OPT2 (LRF 8C)
ABEmax (KLF1)
Cas9 (-197)

Figure 7

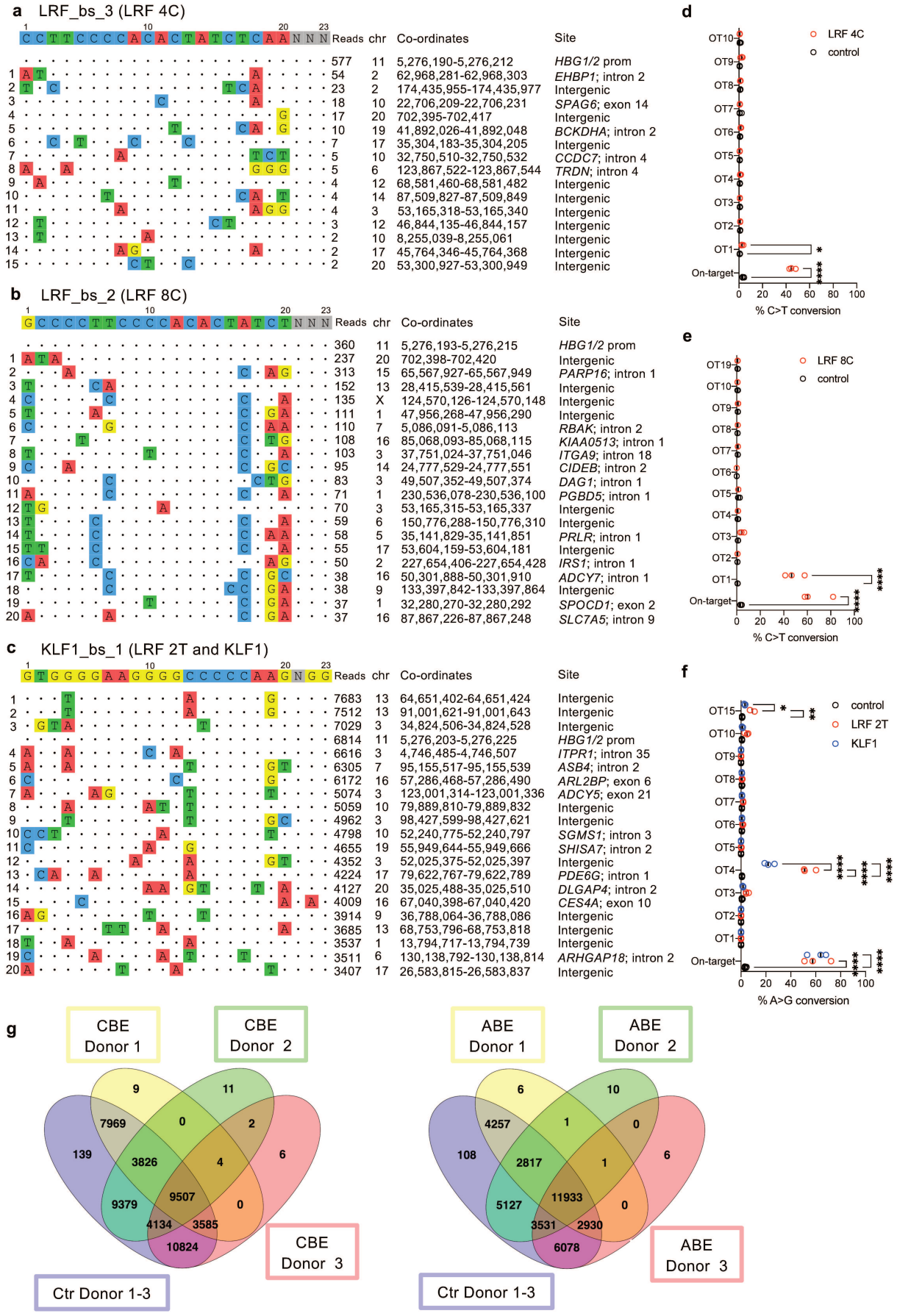
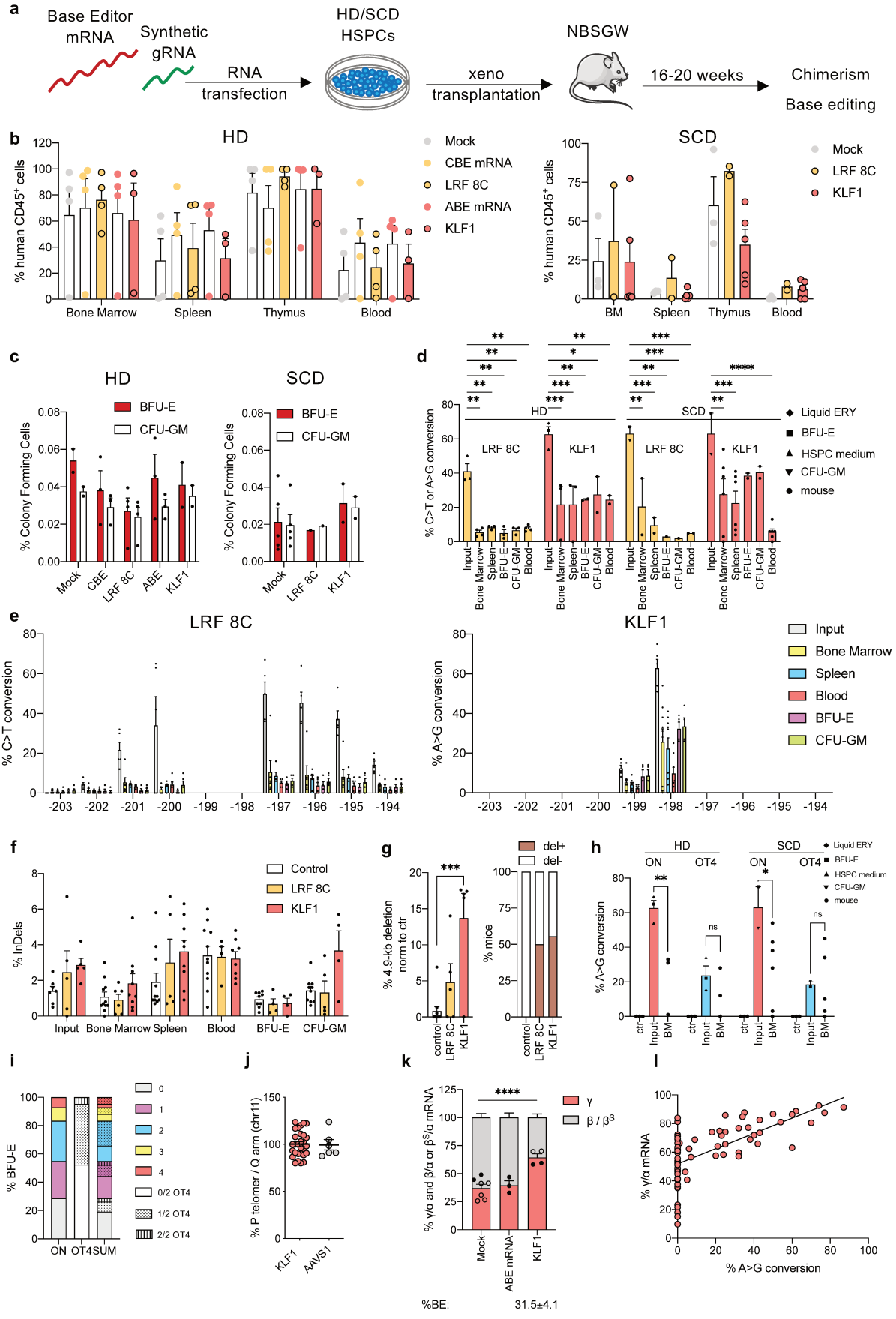

a LRF_bs_3 (LRF 4C)
Reads chr Co-ordinates Site
577 11 5,276,190-5,276,212 HBG1/2 prom
54 2 62,968,281-62,968,303 EHBP1; intron 2
23 2 174,435,955-174,435,977 Intergenic
18 10 22,706,209-22,706,231 SPAG6; exon 14
17 20 702,395-702,417 Intergenic
10 19 41,892,026-41,892,048 BCKDHA; intron 2
7 17 35,304,183-35,304,205 Intergenic
5 10 32,750,510-32,750,532 CCDC7; intron 4
5 6 123,867,522-123,867,544 TRDN; intron 4
4 12 68,581,460-68,581,482 Intergenic
4 14 87,509,827-87,509,849 Intergenic
4 3 53,165,318-53,165,340 Intergenic
3 12 46,844,135-46,844,157 Intergenic
2 10 8,255,039-8,255,061 Intergenic
2 17 45,764,346-45,764,368 Intergenic
2 20 53,300,927-53,300,949 Intergenic
b LRF_bs_2 (LRF 8C)
Reads chr Co-ordinates Site
360 11 5,276,193-5,276,215 HBG1/2 prom
237 20 702,398-702,420 Intergenic
313 15 65,567,927-65,567,949 PARP16; intron 1
152 13 28,415,539-28,415,561 Intergenic
135 X 124,570,126-124,570,148 Intergenic
111 1 47,956,268-47,956,290 Intergenic
110 7 5,086,091-5,086,113 RBAK; intron 2
108 16 85,068,093-85,068,115 KIAA0513; intron 1
103 3 37,751,024-37,751,046 ITGA9; intron 18
95 14 24,777,529-24,777,551 CIDEB; intron 2
83 3 49,507,352-49,507,374 DAG1; intron 1
71 1 230,536,078-230,536,100 PGBD5; intron 1
70 3 53,165,315-53,165,337 Intergenic
59 6 150,776,288-150,776,310 Intergenic
58 5 35,141,829-35,141,851 PRLR; intron 1
55 17 53,604,159-53,604,181 Intergenic
50 2 227,654,406-227,654,428 IRS1; intron 1
38 16 50,301,888-50,301,910 ADCY7; intron 1
38 9 133,397,842-133,397,864 Intergenic
37 1 32,280,270-32,280,292 SPOCD1; exon 2
37 16 87,867,226-87,867,248 SLC7A5; intron 9
c KLF1_bs_1 (LRF 2T and KLF1)
Reads chr Co-ordinates Site
7683 13 64,651,402-64,651,424 Intergenic
7512 13 91,001,621-91,001,643 Intergenic
7029 3 34,824,506-34,824,528 Intergenic
6814 11 5,276,203-5,276,225 HBG1/2 prom
6616 3 4,746,485-4,746,507 ITPR1; intron 35
6305 7 95,155,517-95,155,539 ASB4; intron 2
6172 16 57,286,468-57,286,490 ARL2BP; exon 6
5074 3 123,001,314-123,001,336 ADCY5; exon 21
5059 10 79,889,810-79,889,832 Intergenic
4962 3 98,427,599-98,427,621 Intergenic
4798 10 52,240,775-52,240,797 SGMS1; intron 3
4655 19 55,949,644-55,949,666 SHISA7; intron 2
4352 3 52,025,375-52,025,397 Intergenic
4224 17 79,622,767-79,622,789 PDE6G; intron 1
4127 20 35,025,488-35,025,510 DLGAP4; intron 2
4009 16 67,040,398-67,040,420 CES4A; exon 10
3914 9 36,788,064-36,788,086 Intergenic
3685 13 68,753,796-68,753,818 Intergenic
3537 1 13,794,717-13,794,739 Intergenic
3511 6 130,138,792-130,138,814 ARHGAP18; intron 2
3407 17 26,583,815-26,583,837 Intergenic
d
LRF 4C
control
OT10
OT9
OT8
OT7
OT6
OT5
OT4
OT3
OT2
OT1
On-target
% C>T conversion
e
LRF 8C
control
% C>T conversion
f
control
LRF 2T
KLF1
% A>G conversion
g
CBE Donor 1
CBE Donor 2
CBE Donor 3
Ctr Donor 1-3
9
11
7969
0
2
139
3826
4
6
9379
9507
0
4134
3585
10824
ABE Donor 1
ABE Donor 2
ABE Donor 3
Ctr Donor 1-3
6
10
4257
1
0
108
2817
1
6
5127
11933
0
3531
2930
6078

Figure 8


a
Base Editor mRNA
Synthetic gRNA
RNA transfection
HD/SCD HSPCs
xeno transplantation
NBSGW
16-20 weeks
Chimerism
Base editing
b
HD
% human CD45+ cells
Bone Marrow
Spleen
Thymus
Blood
Mock
CBE mRNA
LRF 8C
ABE mRNA
KLF1
SCD
BM
c
HD
SCD
BFU-E
CFU-GM
% Colony Forming Cells
Mock
CBE
LRF 8C
ABE
KLF1
d
% C>T or A>G conversion
Liquid ERY
BFU-E
HSPC medium
CFU-GM
mouse
e
LRF 8C
KLF1
% C>T conversion
% A>G conversion
Input
Bone Marrow
Spleen
Blood
BFU-E
CFU-GM
f
% InDels
Control
LRF 8C
KLF1
g
% 4.9-kb deletion norm to ctr
del+
del-
% mice
h
HD
SCD
ON
OT4
% A>G conversion
ctr
i
% BFU-E
ON
OT4
SUM
0/2 OT4
1/2 OT4
2/2 OT4
j
% P telomer / Q arm (chr11)
AAVS1
k
% γ/α and β/α or βS/α mRNA
γ
β / βS
%BE:
31.5±4.1
l
% γ/α mRNA
% A>G conversion

Supplementary Information

Base-editing-mediated dissection of a γ-globin *cis*-regulatory element for the therapeutic reactivation of fetal hemoglobin expression

Panagiotis Antoniou[1], Giulia Hardouin[1,2,3], Pierre Martinucci[1], Giacomo Frati[1], Tristan Felix[1], Anne Chalumeau[1], Letizia Fontana[1], Jeanne Martin[1], Cecile Masson[4], Megane Brusson[1], Giulia Maule[5], Marion Rosello[6], Carine Giovannangeli[7], Vincent Abramowski[8], Jean-Pierre de Villartay[8], Jean-Paul Concordet[7], Filippo Del Bene[6], Wassim El Nemer[9,10,11], Mario Amendola[12,13], Marina Cavazzana[3,14,15], Anna Cereseto[5], Oriana Romano[16], Annarita Miccio[1]*

[1]Université Paris Cité, Imagine Institute, Laboratory of chromatin and gene regulation during development, INSERM UMR 1163, 75015, Paris, France
[2]Université Paris Cité, Imagine Institute, Laboratory of Human Lymphohematopoiesis, INSERM UMR 1163, 75015, Paris, France
[3]Biotherapy Department and Clinical Investigation Center, Assistance Publique Hopitaux de Paris, INSERM, 75015 Paris, France
[4]Bioinformatics Platform, Imagine Institute, 75015, Paris, France.
[5]CIBIO, University of Trento, 38100, Trento, Italy
[6]Sorbonne Université, INSERM, CNRS, Institut de la Vision, 75015, Paris, France
[7]INSERM U1154, CNRS UMR7196, Museum National d'Histoire Naturelle, Paris, France
[8]Université Paris Cité, Imagine Institute, Laboratory of genome dynamics in the immune system, INSERM UMR 1163, 75015, Paris, France
[9]Biologie Intégrée du Globule Rouge UMR_S1134, Inserm, Univ. Paris Diderot, Sorbonne Paris Cité, Univ. de la Réunion, Univ. des Antilles
[10]Établissement Français du Sang, UMR 7268, 13005, Marseille, France
[11]Laboratoire d'Excellence GR-Ex, 75015, Paris, France
[12]Genethon, 91000, Evry, France
[13]Université Paris-Saclay, Univ Evry, Inserm, Genethon, Integrare research unit UMR_S951, 91000, Evry, France
[14]Université Paris Cité, 75015, Paris, France
[15]Imagine Institute, 75015, Paris, France
[16] Department of Life Sciences, University of Modena and Reggio Emilia, 41125, Modena, Italy

*To whom correspondence should be addressed: A.M., Imagine Institute, 24, Boulevard du Montparnasse, 75015 Paris, France. E-mail address: annarita.miccio@institutimagine.org.

Keywords: base editing, hemoglobinopathies, fetal hemoglobin reactivation, hereditary persistence of fetal hemoglobin

Table of Contents

Supplementary Tables 198

Supplementary Table 1. sgRNA target sequences. 198

Supplementary Table 2. Primers used to detect base-editing and InDels events. 198

Supplementary Table 3. Primers used for ddPCR. 198

Supplementary Table 4. Primers used for targeted deep sequencing. 198

Supplementary Table 5. Primers used for ChIP-qPCR. 200

Supplementary Table 6. Primers used for RT-qPCR. 200

Supplementary notes 202

Supplementary note 1. Selection of BE/sgRNA combinations targeting the -200 region of the *HBG* promoters 202

Supplementary Note 2. Evaluation of InDel in ABE8e-treated samples 203

Supplementary Figures 204

Supplementary Figure 1. LRF BS disruption and KLF1 BS creation in the *HBG1/2* promoters in K562 cell line. 205

Supplementary Figure 2. HbF reactivation upon LRF BS disruption and KLF1 BS creation in the *HBG1/2* promoters in HUDEP-2 cell line. 207

Supplementary Figure 3. Evaluation of the presence of real InDel events in samples bearing the LRF 2T profile. 208

Supplementary Figure 4. Targeted NGS sequencing of the *HBG1/2* promoters in erythroblasts derived from SCD HSPCs transfected with BE- and sgRNA- expressing plasmids. 210

Supplementary Figure 5. Erythroid differentiation upon plasmid-mediated delivery of base editors to SCD HSPCs. 211

Supplementary Figure 6. HbF reactivation in SCD patient erythroid cells upon plasmid-mediated delivery of base editors to HSPCs. 213

Supplementary Figure 7. Analysis of BFU-E and CFU-GM colonies derived from plasmid-transfected SCD HSPCs. 215

Supplementary Figure 8. Targeted NGS sequencing of the *HBG1/2* promoters in erythroid BFU-E single colonies derived from SCD HSPCs transfected with ABE8e- and KLF_bs_1 sgRNA-expressing plasmids. 216

Supplementary Figure 9. Plasmid optimization for CBE-SpRY *in vitro* transcription. 217

Supplementary Figure 10. Targeted NGS sequencing of the *HBG1/2* promoters in erythroblasts derived from SCD HSPCs transfected with BE mRNA and chemically modified sgRNAs. 220

Supplementary Figure 11. CFC assay from RNA-transfected SCD HSPCs 222

Supplementary Figure 12. Erythroid differentiation of SCD HSPCs upon RNA-mediated delivery of base editors. 224

Supplementary Figure 13. CFC assay from RNA-transfected β-thalassemic HSPCs. 225

Supplementary Figure 14. Analysis of genes activated by RNA stimuli. 227

Supplementary Figure 15. sgRNA-dependent DNA off-target cleavage activity of CBE or ABE targeting the HBG promoters. 228

Supplementary Figure 16. Human hematopoietic cell reconstitution in NBSGW mice transplanted with control and edited HSPCs. 229

Supplementary Figure 17. Detection of the 4.9-kb deletion in repopulating HSCs. 230

Supplementary Figure 18. Examples of gates used to assess transfection efficiency or to flow sort GFP$^+$ cells. 232
Supplementary Figure 19. Examples of gates used to assess erythroid surface markers and enucleated cells in HSPC-derived erythroid cells. 234
Supplementary Figure 20. Examples of gates used to assess HbF and HbS expression. 236
Supplementary Figure 21. Examples of gates used to assess apoptosis and ROS in HSPC-derived erythroid cells. 238
Supplementary Figure 22. Examples of gates used to assess chimerism and lineage specific markers in xenotransplantation experiments. 239
Supplementary sequences 240
CBE-SpRY-OPT sequence 240
Supplementary uncropped scans of blot and gels 243
Uncropped scan of gel from Supplementary Figure 17 243
Supplementary References 247

Supplementary Tables

Supplementary Table 1. sgRNA target sequences.

sgRNA	Target sequence (5' to 3')	Position (hg19)	Strand
KLF1_bs_1	GTGGGGAAGGGGCCCCCAAG	chr11: 5271279-5271298 (*HBG1*) chr11: 5276203-5276222 (*HBG2*)	+
LRF_bs_1	GGGCCCCTTCCCCACACTAT	chr11: 5271269-5271288 (*HBG1*) chr11: 5276193-5276212 (*HBG2*)	-
LRF_bs_2	GCCCCTTCCCCACACTATCT	chr11: 5271272-5271291 (*HBG1*) chr11: 5276196-5276215 (*HBG2*)	-
LRF_bs_3	CCTTCCCCACACTATCTCAA	chr11: 5271274-5271293 (*HBG1*) chr11: 5276198-5276217 (*HBG2*)	-
LRF_bs_4	CTTCCCCACACTATCTCAAT	chr11: 5271268-5271287 (*HBG1*) chr11: 5276192-5276211 (*HBG2*)	-
LRF_bs_5	TTCCCCACACTATCTCAATG	chr11: 5271267-5271286 (*HBG1*) chr11: 5276191-5276210 (*HBG2*)	-
AAVS1	GGGGCCACTAGGGACAGGAT	chr19: 55627120-55627139	-

Supplementary Table 2. Primers used to detect base-editing and InDels events.

Amplified region	F/R	Sequence (5' to 3')
HBG1 + *HBG2* promoters	F	AAAAACGGCTGACAAAAGAAGTCCTGGTAT
	R	ATAACCTCAGACGTTCCAGAAGCGAGTGTG
AAVS1 site	F	CAGCACCAGGATCAGTGAAA
	R	CTATGTCCACTTCAGGACAGCA
4.9-kb deletion	F	GTTTTAAAACAACAAAAATGAGGGAAAGA
	R	GTTGCTTTATAGGATTTTTCACTACAC

F, forward primer; R, reverse primer.

Supplementary Table 3. Primers used for ddPCR.

Amplified region	F/R	Sequence (5' to 3')
HBG1-HBG2 intervening region	F	ACGGATAAGTAGATATTGAGGTAAGC
	R	GTCTCTTTCAGTTAGCAGTGG
hALB	F	ACTCATGGGAGCTGCTGGTT
	R	GCTGTCATCTCTTGTGGGCTG
CARS	F	GGGCCAGGGAAGTGTATGATG
	R	ACAGACATCAGTGCCATTGCG
PODL1	F	GCAGGTTCAGTCCCTCTTGG
	R	TGCTTGGCCTATGGACAGTTG

F, forward primer; R, reverse primer.

Supplementary Table 4. Primers used for targeted deep sequencing.

Amplified region	F/R	Sequence (5' to 3')
HBG promoters	F	GGAATGACTGAATCGGAACAAGG
	R	CTGGCCTCACTGGATACTCT
g3-OT1	F	GCAGCGTCAGATGTGTATAAGAGACAGTCTGTGTGGTCACTCAGGGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCACCTCTAGAACATGAGAAGGGG
g3-OT2	F	GCAGCGTCAGATGTGTATAAGAGACAGGAGGTTGGTAAGAGCAGCGC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTACCCCTTTCTCTAGCAAGTCAG
g3-OT3	F	GCAGCGTCAGATGTGTATAAGAGACAGGGGAGAGCATGTATGCAGGAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTCTGTGGGTAACCTGTTAGGTTTTACTAAG
g3-OT4	F	GCAGCGTCAGATGTGTATAAGAGACAGACAGAAACAAAGGCTTGGAGGTGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCATGGTTCAAGTCCAGAGCTTTCC

g3-OT5	F	GCAGCGTCAGATGTGTATAAGAGACAGATCAGGCATCAGTGCCTTCACAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTGCTCTCCTGCTCTAGCTCG
g3-OT6	F	GCAGCGTCAGATGTGTATAAGAGACAGAGGTTGAAACTCCTCGCCACAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGAGAGTAGTTGAGGCAAGGGAC
g3-OT7	F	GCAGCGTCAGATGTGTATAAGAGACAGAGACCGTTTAGCAGGAACAGA
	R	TGGGCTCGGAGATGTGTATAAGAGACAGATTTAAATTCCTTTCAGAGGAGCAT
g3-OT8	F	GCAGCGTCAGATGTGTATAAGAGACAGATGGGATCCAAACTTACAGTTTCAGCCTTC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTGGATGCCTTTGCCATAGTTGAG
g3-OT9	F	GCAGCGTCAGATGTGTATAAGAGACAGTAGCCAATTCTTCCCATTGGGAGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCCATTGGAATTTAATGACAGCTAGAGTGGG
g3-OT10	F	GCAGCGTCAGATGTGTATAAGAGACAGGTGACACCAAAGCAGATGTCC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGACATAAAGACCCCTCTTGCCT
g12-OT1	F	GCAGCGTCAGATGTGTATAAGAGACAGGAGAGTGAGCAGCAGCAAGTATTG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGACCACTCAGCTCTACCAATCAG
g12-OT2	F	GCAGCGTCAGATGTGTATAAGAGACAGTAAAAGCAGGCTGCCTGAGCCG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCCTTCATGGTGAGTGTTACAGCTG
g12-OT3	F	GCAGCGTCAGATGTGTATAAGAGACAGCATGGAAGAAAGAGAGGGAAGGAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGTGGAATGCCTTCCCTTACTGATC
g12-OT4	F	GCAGCGTCAGATGTGTATAAGAGACAGGAGAGTTGATGTCTGTGGAACGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCGTCACAGCTAGAAACTCTATGCC
g12-OT5	F	GCAGCGTCAGATGTGTATAAGAGACAGGAACAGTGATAAGGAATTTCGAAGCCAGTC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGTGCCAGGAAGAAGTCATTGCTTG
g12-OT6	F	GCAGCGTCAGATGTGTATAAGAGACAGACCTCAAGTAGACACCTCTCTCAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGCAGAGCCTGCTGGAACAGA
g12-OT7	F	GCAGCGTCAGATGTGTATAAGAGACAGTTAGAGGGAACGAAGAGCTAGCAC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGGTTTTCGTTCTGCTTCCAACCTG
g12-OT8	F	GCAGCGTCAGATGTGTATAAGAGACAGAGAAGCTTCCTGAAGTCCAGTTCC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGGGAAAAAGCTCCAGGAAGCTATG
g12-OT9	F	GCAGCGTCAGATGTGTATAAGAGACAGTCTGCAGGCACGTATTTCCCG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGATGGAGACACGGACCAAGAG
g12-OT10	F	GCAGCGTCAGATGTGTATAAGAGACAGTGTGGGAATTTCAGCAACCCCAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTCCATCCAGACTGCCAACAATC
g12-OT15	F	GCAGCGTCAGATGTGTATAAGAGACAGTCTGCATCCAGAGGTGGGAG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTTTGAGCACCACGCTCGTGGA
g21-OT1	F	GCAGCGTCAGATGTGTATAAGAGACAGACAGAAACAAAGGCTTGGAGGTGG
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCATGGTTCAAGTCCAGAGCTTTCC
g21-OT2	F	GCAGCGTCAGATGTGTATAAGAGACAGGTACTCTAATGTTAGTGTGACTGCCATGAC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGTACAGGGTCCTTCACAGGATC
g21-OT3	F	GCAGCGTCAGATGTGTATAAGAGACAGGCTCAGTGGGGATGAAGGGT
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCACACACCCTCATAGGATTCTTTGTCTTC
g21-OT4	F	GCAGCGTCAGATGTGTATAAGAGACAGTGATCTCATCGTACCCAGGTATCC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCTGGGTTTAATGCTGTCTGAAGGG
g21-OT5	F	GCAGCGTCAGATGTGTATAAGAGACAGAAGCCATAGCAAACTACTGAAGTAGGAACC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCCAGATTTAGCCTCACAGGCC
g21-OT6	F	GCAGCGTCAGATGTGTATAAGAGACAGAGGTCTTGGGTTTTCGGGCC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTTCCCATGCAGCGCCTGTT
g21-OT7	F	GCAGCGTCAGATGTGTATAAGAGACAGGGGCAAACCCTACCGAGTGA
	R	TGGGCTCGGAGATGTGTATAAGAGACAGGACTGTACTCCTGCCAAGGAGTT
g21-OT8	F	GCAGCGTCAGATGTGTATAAGAGACAGGCAAGGGCATAAAAATGGGGGAC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTCAGTCACACCCGCCCATCT
g21-OT9	F	GCAGCGTCAGATGTGTATAAGAGACAGGTTTATGAGTGGCTCTGTGTGTGC
	R	TGGGCTCGGAGATGTGTATAAGAGACAGTCCTTACACCCTTCCTGGACAC

g21-OT10	F	GCAGCGTCAGATGTGTATAAGAGACAGTTAGAGGCCCAAAGGCCTGA
	R	TGGGCTCGGAGATGTGTATAAGAGACAGCCATTATGGAAGCCTCGCCC
g21-OT19	F	GCAGCGTCAGATGTGTATAAGAGACAGACAGGAGGTTGTCCCCACGA
	R	TGGGCTCGGAGATGTGTATAAGAGACAGAAAGAGGAGCCCCCTGGAAG
UDI-1	F	AATGATACGGCGACCACCGAGATCTACACAATAACGTTCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATTAACGATTGTCTCGTGGGCTCGGAGATGT
UDI-2	F	AATGATACGGCGACCACCGAGATCTACACTTCTTGAATCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATATGTAGACGTCTCGTGGGCTCGGAGATGT
UDI-3	F	AATGATACGGCGACCACCGAGATCTACACGGCAGATCTCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATGAGCAGCGGTCTCGTGGGCTCGGAGATGT
UDI-4	F	AATGATACGGCGACCACCGAGATCTACACCTATGTTATCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATTGTTGATCGTCTCGTGGGCTCGGAGATGT
UDI-5	F	AATGATACGGCGACCACCGAGATCTACACGTTGACGCTCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATGTCCTTCGGTCTCGTGGGCTCGGAGATGT
UDI-6	F	AATGATACGGCGACCACCGAGATCTACACATCTACGATCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATCCGGCATCGTCTCGTGGGCTCGGAGATGT
UDI-7	F	AATGATACGGCGACCACCGAGATCTACACCTCGACAGTCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATCTTCGTAGGTCTCGTGGGCTCGGAGATGT
UDI-8	F	AATGATACGGCGACCACCGAGATCTACACGAGGCTGCTCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATGACGCATCGTCTCGTGGGCTCGGAGATGT
UDI-9	F	AATGATACGGCGACCACCGAGATCTACACCCTCGTAGTCGTCGGCAGCGTCAGATGTG
	R	CAAGCAGAAGACGGCATACGAGATTGCCGTAGGTCTCGTGGGCTCGGAGATGT

F, forward primer; R, reverse primer.

Supplementary Table 5. Primers used for ChIP-qPCR.

Amplified region	F/R	Sequence (5' to 3')
HBG	F	TCAATGCAAATATCTGTCTGAAACG
	R	CAAGGCTATTGGTCAAGGCAA
KLF1	F	CCCAACCCAGGCAAATTG
	R	GGGCTGGGAGTTGGGTCTT
DEFB122	F	TGTGGCTGGTCCTTGGGCTT
	R	GTGGCTCCTGCCGTGACGAA

Supplementary Table 6. Primers used for RT-qPCR.

Amplified region	F/R	Sequence (5' to 3')
HBA	F	CGGTCAACTTCAAGCTCCTAA
	R	ACAGAAGCCAGGAACTTGTC
HBB	F	GCAAGGTGAACGTGGATGAAGT
	R	TAACAGCATCAGGAGTGGACAGA
HBG1+HBG2	F	CCTGTCCTCTGCCTCTGCC
	R	GGATTGCCAAAACGGTCAC
HBD	F	CAAGGGCACTTTTTCTCAG
	R	AATTCCTTGCCAAAGTTGC
HBE	F	CTTTGGAAACCTGTCGTC
	R	CTTGCCAAAGTGAGTAGC
CDKN1	F	CAGCATGACAGATTTCTACCACTC
	R	CTCGCGCTTCCAGGACTG
TNF-α	F	CCAGGGACCTCTCTCTAATCAGC
	R	GGTTTGCTACAACATGGGCTAC
IL-6	F	GATTCAATGAGGAGACTTGCCTGG
	R	CTCACTACTCTCAAATCTGTTCTGG
IL-12	F	ACCACTCCCAAAACCTGC

	R	CCAGGCAACTCCCATTAG
IFN-α	F	GACTCCATCTTGGCTGTGA
	R	TGATTTCTGCTCTGACAACCT
IFN-β	F	TACTGCCTCAAGGACAGGATGAA
	R	GCATCTCATAGATGGTCAATGCG
TLR7	F	CTGACCACTGTCCCTGAG
	R	AACCCACCAGACAAACCA
TLR8	F	AACATCAGCAAGACCCAT
	R	GACTCCTTCATTCTCCCT
RIG-I	F	GGACGTGGCAAAACAAATCAG
	R	GCAATGTCAATGCCTTCATCA
GAPDH	F	GAAGGTGAAGGTCGGAGT
	R	GAAGATGGTGATGGGATTTC

F, forward primer; R, reverse primer.

Supplementary notes

Supplementary note 1. Selection of BE/sgRNA combinations targeting the -200 region of the *HBG* promoters

Initially, we explored CBEs targeting the Cs in the LRF BS. The absence of a canonical SpyCas9 NGG PAM close to the LRF BS prompted us to use CBEs containing non-NGG Cas9 nickases. We designed 5 sgRNAs and screened them with a variety of CBEs in the K562 fetal erythroid cell line by plasmid transfection (**Supplementary Figure 1a-c**). For two sgRNAs (LRF_bs_1 and LRF_bs_2), most of the cytosines of the LRF motif are located within the editing window (positions 3-8). The other three sgRNAs (LRF_bs_3, LRF_bs_4 and LRF_bs_5) place only the second cytosine stretch in the editing window. This allowed the creation of two editing profiles, LRF 8C (up to 8 Cs converted to T) and LRF 4C (4 Cs converted to T) (**Figure 1a, Supplementary Figure 1a-c**).

First, we combined the deaminase of the highly efficient AncBE4max enzyme[1] and the SaKKH-Cas9 nickase recognizing an NNNRRT PAM[2], to generate SaKKH-AncBE4max and we targeted the LRF BS with LRF_bs_1 or LRF_bs_4 sgRNAs. Sanger sequencing showed poor editing efficiency (**Supplementary Figure 1a-c**).

We then exchanged the PAM-interacting domain (PID) of the SpyCas9 nickase of AncBE4max with the PID of S.macacae-Cas9 recognizing an NAA PAM. AncBE4max-NAA[3] in combination with LRF_bs_5 or LRF_bs_2 sgRNA was able to modify 4 or 7 out of the 8 cytosines with efficiencies of 25.8%±5.0 when using LRF_bs_2 sgRNA (top efficiency, i.e., the highest base conversion efficiency at the target site; **Supplementary Figure 1a-c**). Codon optimization of S.macacae-Cas9 PID modestly increased the editing efficiency (**Supplementary Figure 1c**).

Furthermore, evoFERNY-BE4max-NG or evoCDA1-BE4max-NG CBEs recognizing the more flexible NG PAM[4], combined with LRF_bs_3 sgRNA, were able to target 5 or 7 out of 8 cytosines of the LRF motif, with the highest but variable frequency observed with evoCDA1-BE4max-NG (47.0%±19.6; **Supplementary Figure 1a-c**).

The best performing base editors were CBEs recognizing alternative and/or more flexible PAMs (i.e., CBE-NRCH, CBE-NRRH, CBE-SpG compatible with NG PAM, and the PAMless CBE-SpRY)[5,6]. CBE-NRCH/LRF_bs_3 sgRNA, CBE-NRRH/LRF_bs_2 sgRNA and CBE-SpG/ LRF_bs_3 sgRNA combinations resulted in efficiencies of 50.3%±4.7, 44.0%±5.3 and 43.7%±3.9, respectively, as evaluated by Sanger sequencing (**Supplementary Figure 1a-c**). CBE-SpRY enzyme allowed more combinations with sgRNAs (LRF_bs_1, LRF_bs_2, LRF_bs_3 and LRF_bs_5) and hit efficiencies of up to 58.0%±1.5 using LRF_bs_2 sgRNA (**Supplementary Figure 1a-c**).

Overall, we selected 5 combinations of CBEs and sgRNAs (CBE-NRCH/LRF_bs_3, CBE-NRRH/LRF_bs_2, CBE-SpG/LRF_bs_3, CBE-SpRY/LRF_bs_3 and CBE-SpRY/LRF_bs_2), which were associated with the highest efficiency and different editing profiles (three generating the LRF 4C profile and two generating the LRF 8C profile).

Next, we used ABEmax and the KLF1_bs_1 sgRNA to generate a KLF1 BS with an efficiency of 30.3% ± 3.8 (KLF1 profile; **Figure 1a, Supplementary Figure 1a and d**). In parallel, we used the ABE8e enzyme containing a highly processive deaminase[7] with the same sgRNA, resulting in efficient conversion of both central Ts (72.7% ± 1.5 for A_7; 66.0% ± 1.5 for A_8) (LRF 2T profile; **Figure 1a, Supplementary Figure 1a and d**).

As fetal K562 cells express high HbF levels, we employed the HUDEP-2 adult erythroid cell line (producing mainly β-globin) to evaluate γ-globin de-repression upon plasmid delivery of the base-editing system (**Supplementary Figure 2a**). We targeted the -200 region of the *HBG* promoters with CBEs to disrupt the LRF BS, achieving an overall low editing frequency, even with enzymes that were highly efficient in K562 (e.g., CBE-NRRH), and low HbF levels (**Supplementary Figure 2b**). On the contrary, samples treated with ABEmax or ABE8e showed high base-editing efficiency and frequency of HbF-expressing cells (**Supplementary Figure 2c**).

Lastly, to confirm that base editors induce little or no DSBs compared to Cas9 nuclease[8], we measured the frequency of InDels in base-edited K562 and HUDEP-2 samples by Sanger sequencing. In most of the cases, we detected no InDels, except for evoCDA1-BE4max-NG[4,9] , while for the highly processive ABE8e enzyme 1 or 2-nt InDels did not occur at the expected

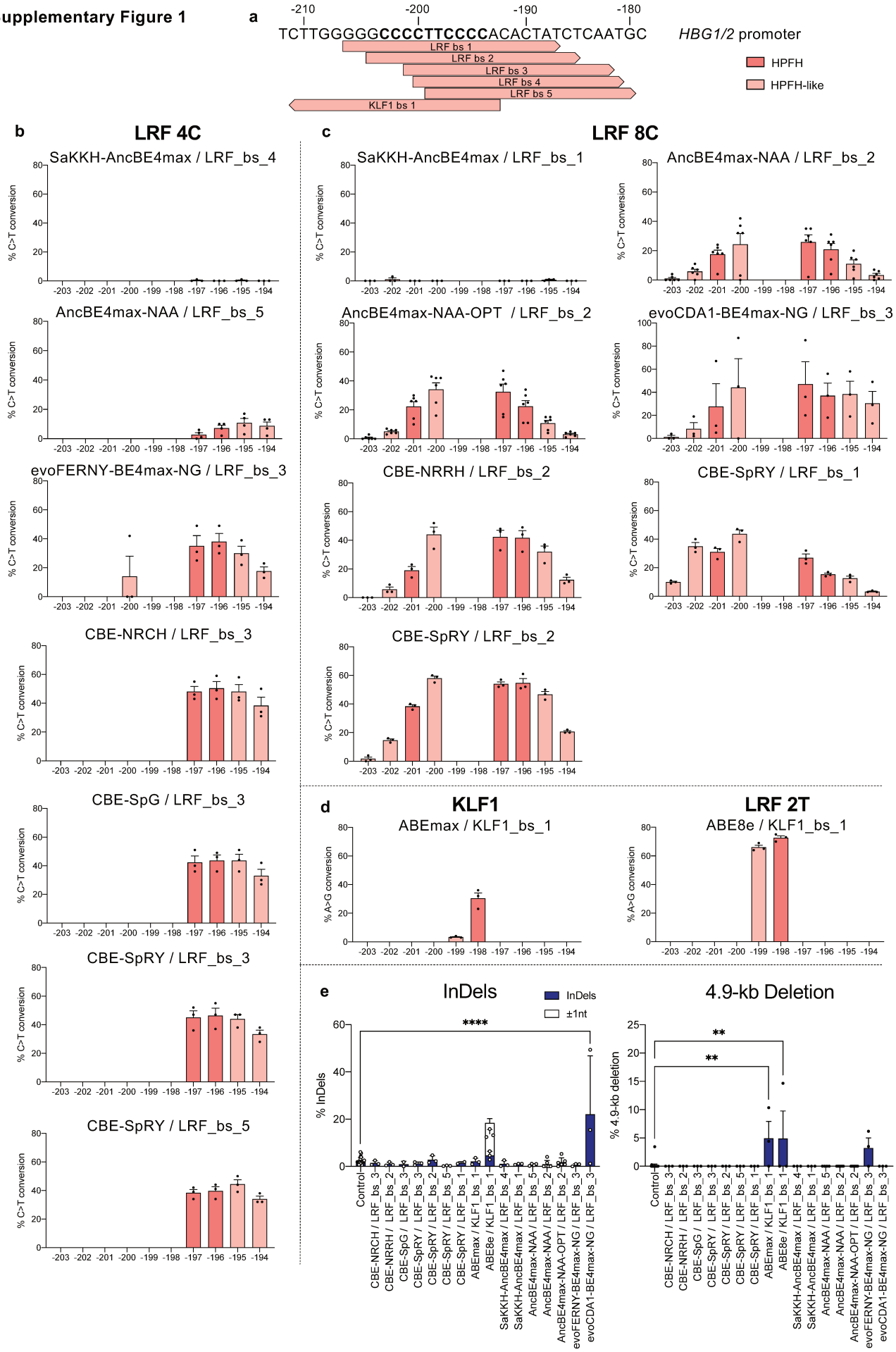
cleavage site and are likely observed because of either a polymerase or a sequencing error due to the presence of a 10-nt-long C homopolymer in edited *HBG* promoters (**Figure 1a, Supplementary Figure 1e, Supplementary Figure 2d, Supplementary Note 2**). Furthermore, we detected a low frequency of the 4.9-kb deletion (resulting from the simultaneous cleavage of the identical *HBG1/2* promoters) only for some BEs (evoFERNY-BE4max-NG, ABEmax and ABE8e) (**Supplementary Figure 1e and Supplementary Figure 2d**).

Supplementary Note 2. Evaluation of InDel in ABE8e-treated samples

The evaluation of InDel frequency by TIDE[10] analysis in K562 edited samples subjected to Sanger sequencing, revealed ~5-15% of InDels in ABE8e-treated samples bearing the LRF 2T profile. As TIDE[11] analysis does not provide the precise location of the InDels, we used ICE Synthego. This analysis confirmed the presence of InDels at the expected cleavage site, with similar frequencies. Initially, and being encouraged/biased by the first report of ABE8e enzyme that revealed a higher tendency of ABE8e to cause DSBs, as compared to ABEmax[12], we concluded that these events are real InDel events. These results were confirmed in ABE8e-treated SCD HSPCs. Nevertheless, ICE is not accurate in identifying the position of InDels, as the possible genotypes encompass all frequent InDels identified in the literature (e.g. typically +1 and -1 InDels at the expected cleavage site located 3-4 nucleotides before the PAM). Indeed, Sanger sequencing chromatograms in edited samples showed the presence of the 1 nucleotide deletion in the poly-C homopolymers and not at the expected cleavage site (**Supplementary Figure 3a**). Similarly, NGS sequencing and CRISPRESSO 2 analysis (more precise as compared to Sanger sequencing and TIDE/ICE analyses) of ABE8e-treated SCD HSPCs revealed high frequencies of InDels, that, however, were not present at the expected cleavage site. In particular, NGS revealed the presence of 1 or 2 nucleotide deletions, and with a lower frequency the presence of 1 C insertion, within the poly-C homopolymer stretch that was present in ABE8e treated samples (LRF 2T profile) (**Supplementary Figure 4d**). The frequency of these types of InDels was higher in samples with high base-editing efficiency and consequently high presence of the homopolymer C-stretch (**Supplementary Figure 3b**). In ABEmax-treated samples (KLF1 profile), when the LRF 2T profile was occasionally present upon modification of both central Ts, there was a small fraction of promoters harboring 1 nucleotide deletion or insertion (**Supplementary Figure 4d**). Based on these considerations, we concluded that the 1-2 nucleotide deletions or 1 nucleotide insertion observed in LRF 2T, and sometimes in KLF1 samples, are likely observed either because of a polymerase error or because of a sequencing error due to the presence of a 10-nt-long C homopolymer in edited *HBG* promoters.

Supplementary Figures

Supplementary Figure 1



Supplementary Figure 1. LRF BS disruption and KLF1 BS creation in the *HBG1/2* promoters in K562 cell line.

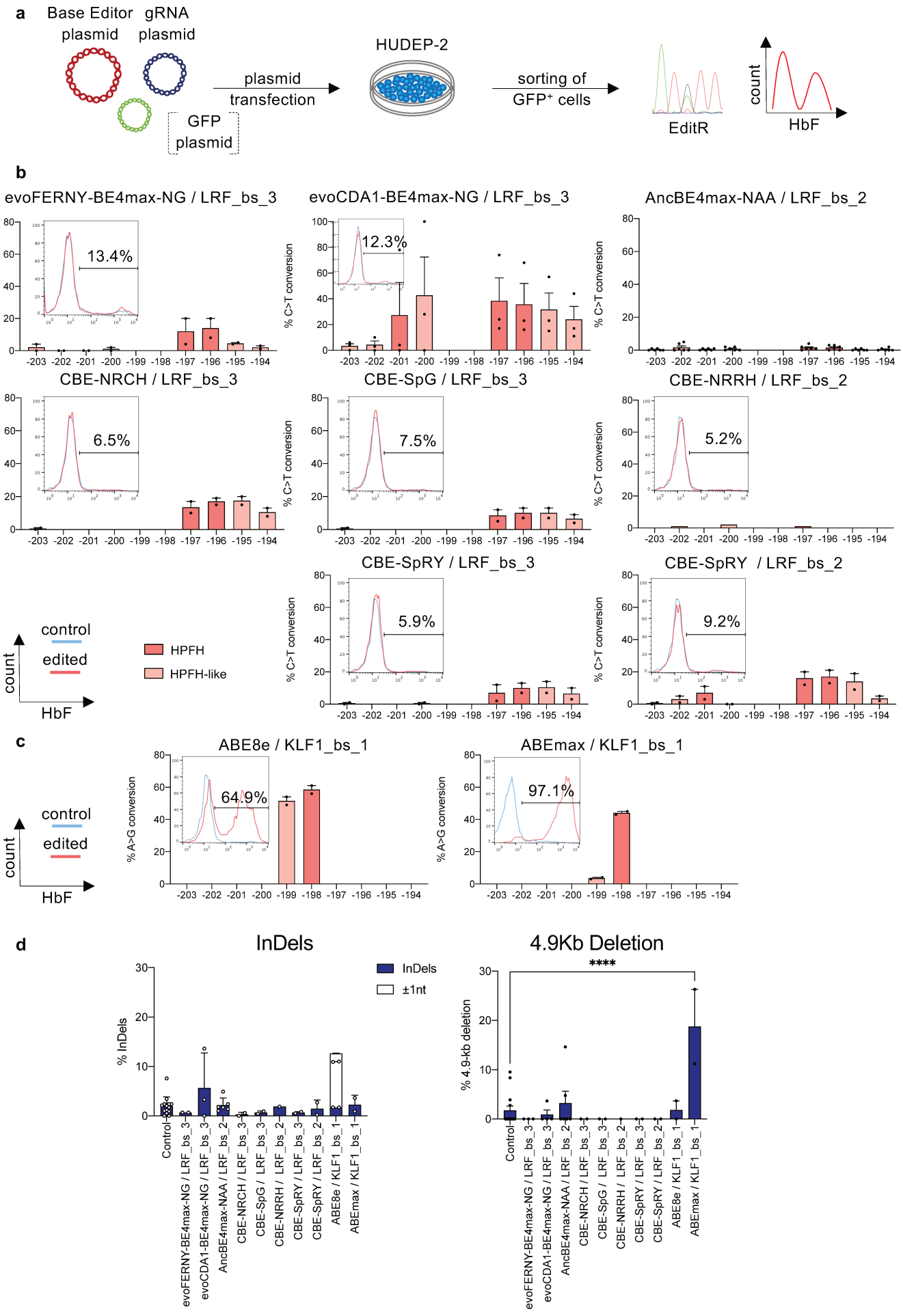
a. Representation of the sequence of the *HBG2* and *HBG1* identical promoters, from –212 to –179 nucleotides upstream of the *HBG* transcription start sites. Pink arrows indicate the sgRNA designed to target the -200 region of the *HBG1* and *HBG2* identical promoters aligned to the target DNA sequences.

b-d. C-G to T-A (b and c) and A-T to G-C (d) base-editing efficiency calculated by the EditR software in samples subjected to Sanger sequencing. The base-editing efficiency was measured by subtracting the percentage of the base conversion in the control that was considered as background noise. On the top of each graph the editing profile, the enzyme and the sgRNA used are indicated. HPFH (dark pink) and HPFH-like (light pink) mutations are indicated. Data are expressed as mean ± SEM [b: n=3 (SaKKH-AncBE4max, evoFERNY-BE4max-NG, CBE-NRCH, CBE-SpG and CBE-SpRY), n=4 (AncBE4max-NAA) biologically independent experiments; c: n=3 (SaKKH-AncBE4max, evoCDA1-BE4max-NG, CBE-NRRH and CBE-SpRY), n=6 (AncBE4max-NAA and AncBE4max-NAA-OPT) biologically independent experiments; d: n=3 biologically independent experiments].

e. Frequency of InDels, measured by TIDE analysis, and frequency of the 4.9-kb deletion, measured by ddPCR, for control (transfected with TE buffer or enzyme plasmid only) and base-edited samples. The insertion or deletion of a C (±1 nt) in the homopoly-C stretch of the LRF 2T profile was separated from the overall frequency of InDels, as it was considered a sequencing error (**Supplementary Note 2**). Data are expressed as mean ± SEM [InDels: n=21 (Control), n=4 (AncBE4max-NAA/LRF_bs_5), n=6 (AncBE4max-NAA/LRF_bs_2 and AncBE4max-NAA-OPT/LRF_bs_2), n=2 (CBE-SpG/LRF_bs_3 and CBE-SpRY/LRF_bs_1) and n=3 (other groups) biologically independent experiments; 4.9-kb deletion: n=19 (Control), n=4 (AncBE4max-NAA/LRF_bs_5), n=6 (AncBE4max-NAA/LRF_bs_2 and AncBE4max-NAA-OPT/LRF_bs_2) and n=3 (other groups) biologically independent experiments]. ** p=0.0054 for ABEmax, or p=0.0060 for ABE8e; **** $p \leq 0.0001$ (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

Source data are provided as a Source Data file.

Supplementary Figure 2



Supplementary Figure 2. HbF reactivation upon LRF BS disruption and KLF1 BS creation in the *HBG1/2* promoters in HUDEP-2 cell line.

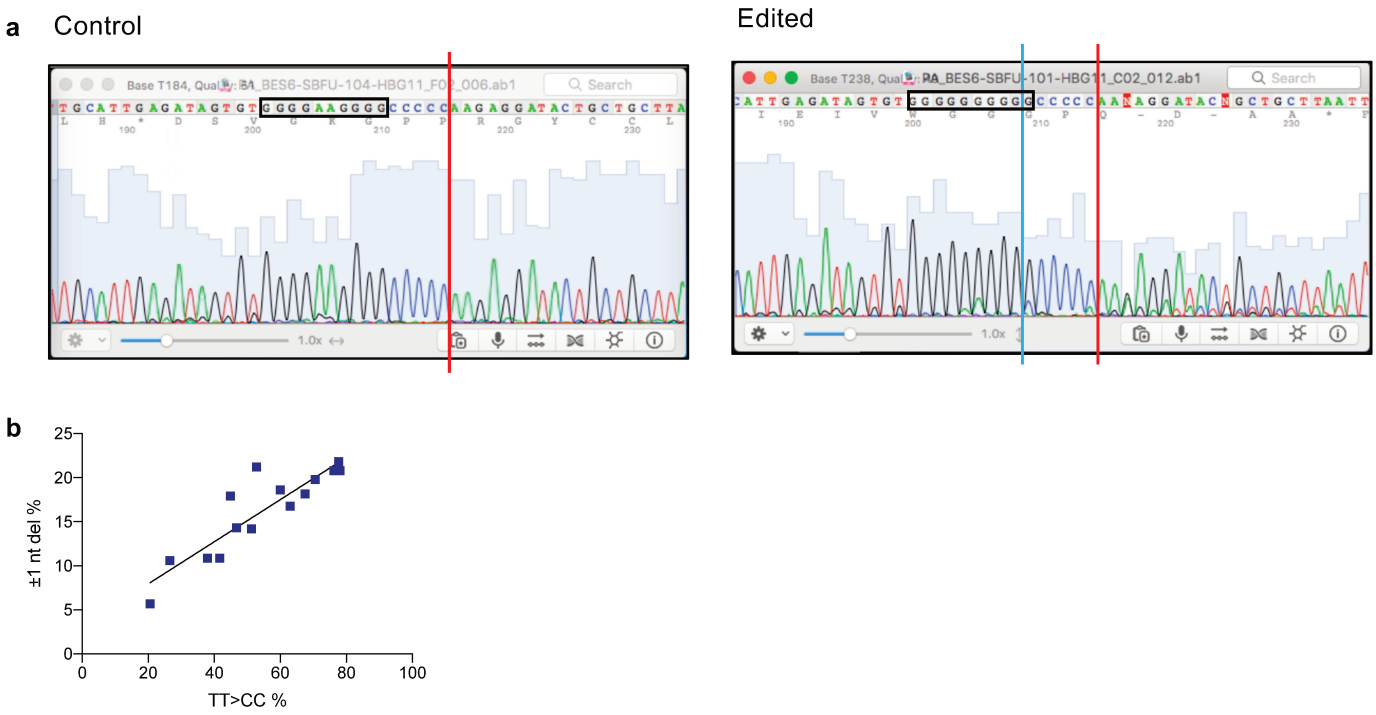
a. Experimental protocol used for base-editing experiments in HUDEP-2 cell line. A BE-, a sgRNA- and a GFP- (optional for enzyme plasmids that do not contain a GFP cassette) expressing plasmids were co-transfected in HUDEP-2 cell line and 18h post-transfection, GFP^+ cells were FACS-sorted and further analyzed for base-editing efficiency and HbF reactivation.

b-c. C-G to T-A (b) and A-T to G-C (c) base-editing efficiency, calculated by the EditR software in samples subjected to Sanger sequencing. On the top of each graph the enzyme and the sgRNA used are indicated. Data are expressed as mean ± SEM [n=2 (evoFERNY-BE4max-NG/LRF_bs_3, CBE-NRCH/LRF_bs_3, CBE-SpG/LRF_bs_3, CBE-SpRY/LRF_bs_3, CBE-SpRY/LRF_bs_2, ABEmax/KLF1_bs_1 and ABE8e/KLF1_bs_1), n=3 (evoCDA1-BE4max-NG/LRF_bs_3), n=6 (AncBE4max-NAA/LRF_bs_2), n=1 (CBE-NRRH/LRF_bs_2) biologically independent experiments]. Representative flow cytometry histograms showing the percentage of HbF-expressing cells for control (transfected with enzyme plasmid only; blue line) and base-edited (red line) samples are reported within each graph.

d. Frequency of InDels, measured by TIDE analysis, and frequency of the 4.9-kb deletion, measured by ddPCR, for control (transfected with TE buffer or enzyme plasmid only) and base-edited samples. The insertion or deletion of a C (±1 nt) in the homopoly-C stretch of the LRF 2T profile was separated from the overall frequency of InDels, as it was considered a sequencing error (**Supplementary Note 2**). Data are expressed as mean ± SEM [InDels: n=14 (Control), n=3 (evoCDA1-BE4max-NG), n=6 (AncBE4max-NAA), n=1 (CBE-NRRH) and n=2 (other groups) biologically independent experiments; 4.9-kb deletion: n=14 (Control), n=3 (evoFERNY-BE4max-NG), n=4 (evoCDA1-BE4max-NG), n=6 (AncBE4max-NAA), n=1 (CBE-NRRH) and n=2 (other groups) biologically independent experiments]. **** $p \leq 0.0001$ (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

Source data are provided as a Source Data file.

Supplementary Figure 3



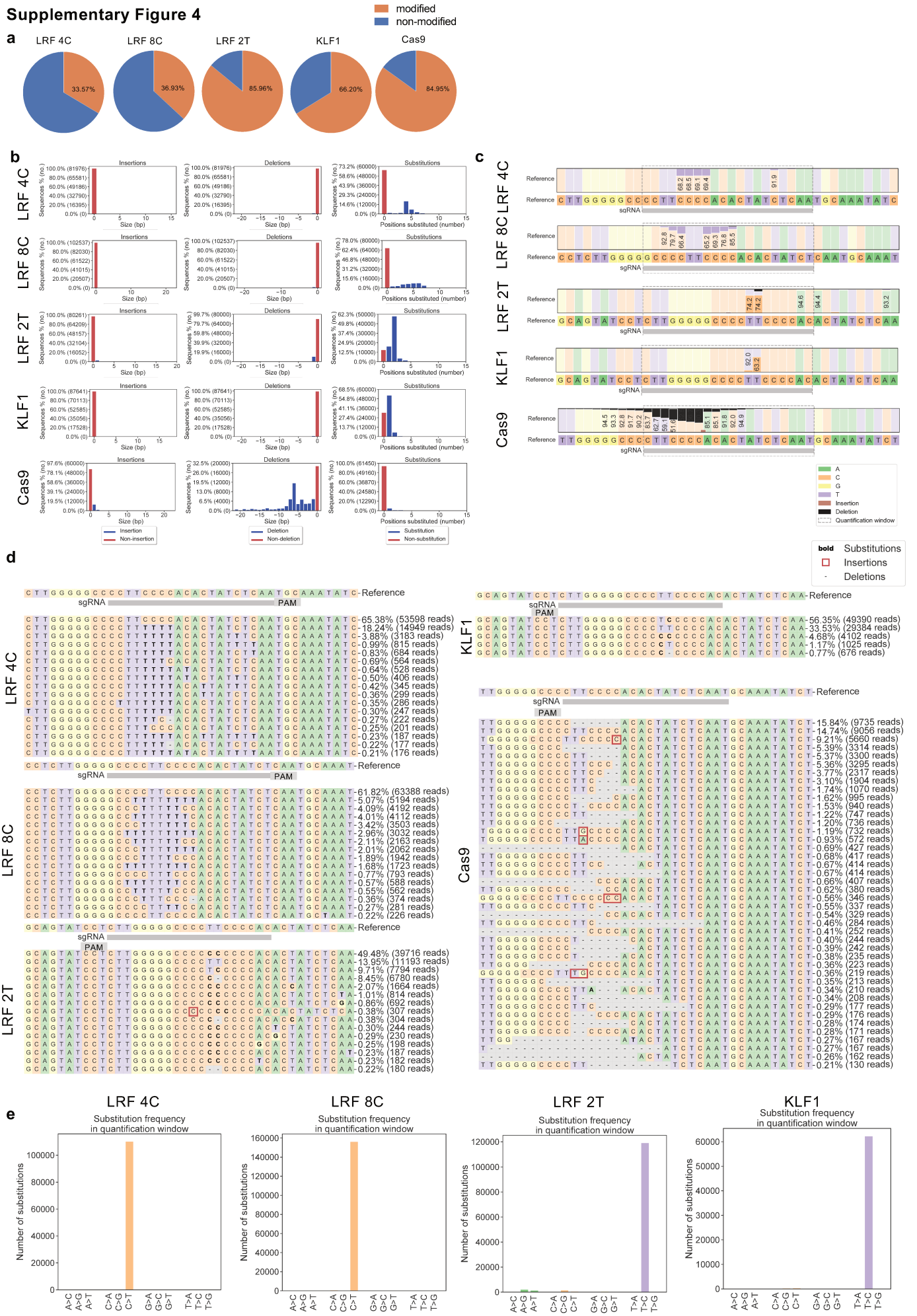
Supplementary Figure 3. Evaluation of the presence of real InDel events in samples bearing the LRF 2T profile.

a. Representative Sanger sequencing chromatograms of control (n=8) and edited BFU-E (n=17; from **Figure 2i**). The red line indicates the expected cleavage site. The blue line indicates the position where the shift of the sequence occurs because of the 1-nt deletion. The black box indicates the LRF BS (in control sample) and the poly-C homopolymer in the edited sample.

b. Correlation between InDels and base editing efficiency. We correlated the frequency of ±1 InDels and base editing efficiency in the two central T of the -200 site (TT>CC) in LRF 2T colonies from **Figure 2i** (n=15 biologically independent BFU-E colonies).

Source data are provided as a Source Data file.

Supplementary Figure 4



Supplementary Figure 4. Targeted NGS sequencing of the *HBG1/2* promoters in erythroblasts derived from SCD HSPCs transfected with BE- and sgRNA- expressing plasmids.

a. Frequency of modified (orange) and unmodified (blue) promoters (reads) in base editor- and Cas9-treated samples sorted for GFP^{high} expression, as measured by targeted NGS sequencing. Data are expressed as single values (1 donor).

b. Frequency of insertions, deletions and substitutions in base editor- and Cas9-treated samples sorted for GFP^{high} expression, as measured by targeted NGS sequencing. Data are expressed as single values (1 donor).

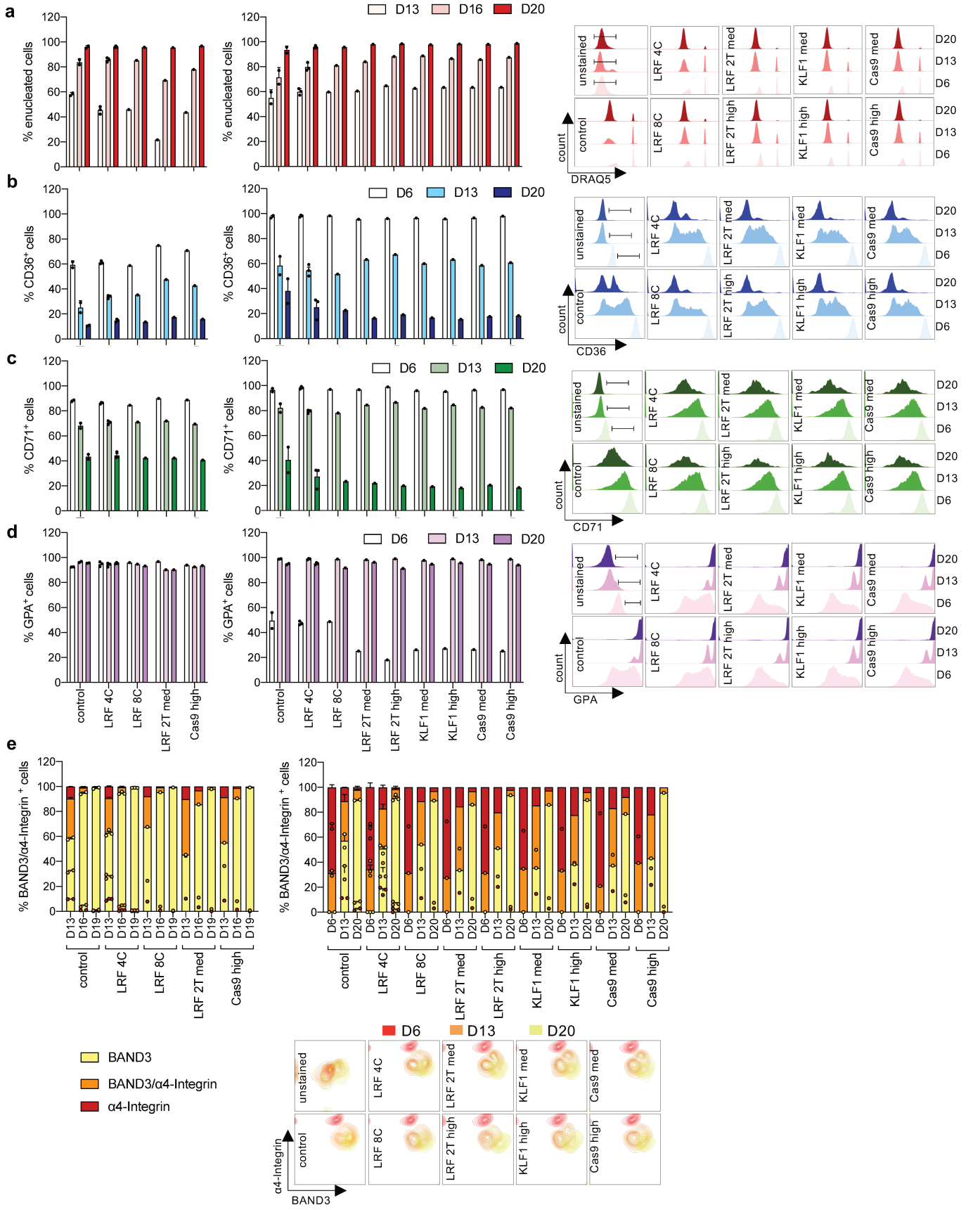
c. Frequency and location of insertions, deletions and substitutions in base editor- and Cas9- treated samples sorted for GFP^{high} expression, as measured by targeted NGS sequencing. Data are expressed as single values (1 donor).

d. Frequency and sequence of modified and unmodified promoters in base editor- and Cas9- treated samples sorted for GFP^{high} expression, as measured by targeted NGS sequencing. Data are expressed as single values (1 donor). In ABE8e-treated samples, around 10% of the total base-editing events were 1-nt or 2-nt deletions, and 1 C insertions. However, the presence of these events is likely due to a polymerase or sequencing error caused by the presence of a 10-nt-long C homopolymer.

e. Product purity of base-editing enzymes, as indicated by the type of substitution and measured by targeted NGS sequencing. Data are expressed as single values (1 donor).

Source data are provided as a Source Data file.

Supplementary Figure 5

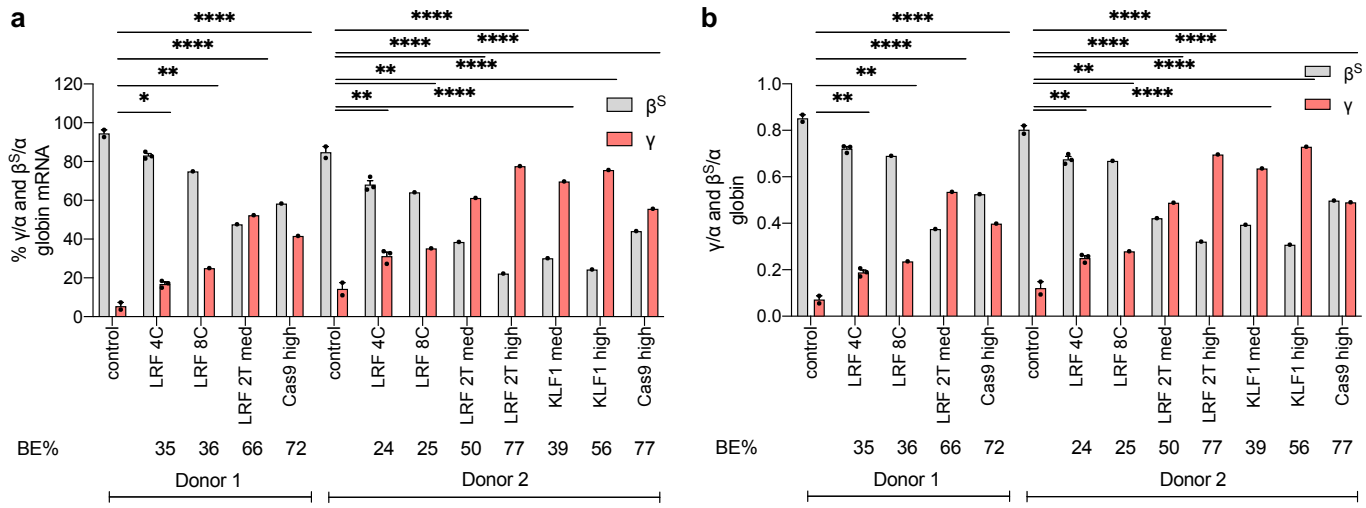


Supplementary Figure 5. Erythroid differentiation upon plasmid-mediated delivery of base editors to SCD HSPCs.

a. Frequency of enucleated cells at day 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of DRAQ5 in control (transfected with TE buffer or with CBE-SpRY plasmid and a sgRNA targeting the unrelated *AAVS1* locus) and edited samples. Data are expressed as

single values or as mean ± SEM [n=2 (control), n=3 (LRF 4C), n=1 (other groups) biologically independent experiments, 2 donors]. Representative flow cytometry histograms showing the $DRAQ5^-$ cell population for unstained, control and edited samples are presented.
b-d. Frequency of $CD36^+$ (b), $CD71^+$ (c) and $GYPA^+$ (d) cells at day 6, 13 and 20 of erythroid differentiation, as measured by flow cytometry analysis of CD36, CD71 and GYPA erythroid markers. Data are expressed as single values or as mean ± SEM [n=2 (control), n=3 (LRF 4C), n=1 (other groups) biologically independent experiments, 2 donors]. Representative flow cytometry histograms showing the $CD36^+$ (b), $CD71^+$ (c) and $GYPA^+$ (d) cell population for unstained, control and edited samples are reported.
e. Frequency of α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ in $7AAD^-$/$GYPA^+$ cells at day 6, 13 and 20 of erythroid differentiation, as measured by flow cytometry analysis of α4-Integrin and BAND3 erythroid markers. Data are expressed as single values or as mean ± SEM [n=2 (control), n=3 (LRF 4C), n=1 (other groups) biologically independent experiments, 2 donors]. Representative flow cytometry contour plots showing the α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ cell population for unstained, control and edited samples are reported.
Source data are provided as a Source Data file.

Supplementary Figure 6



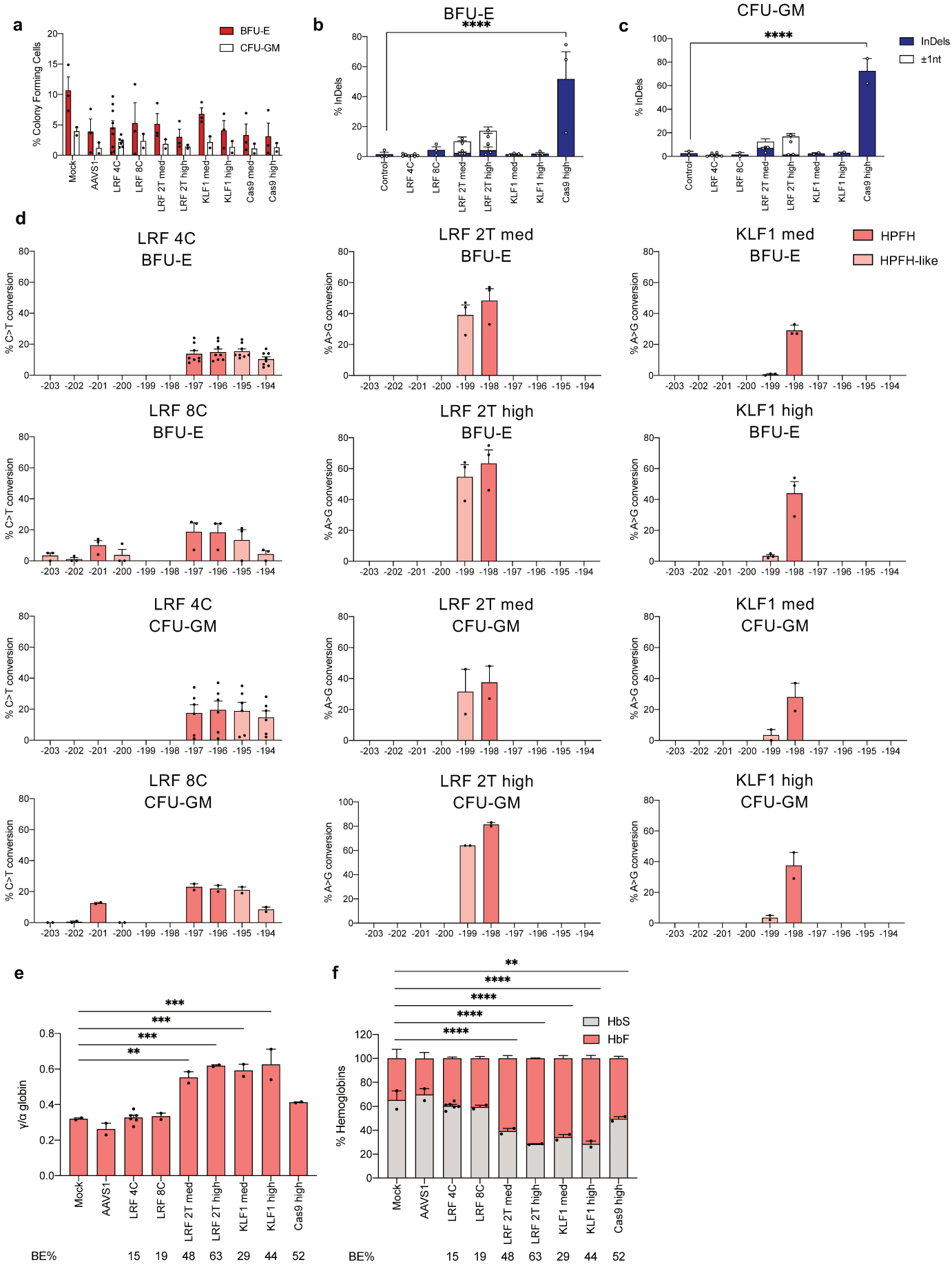
Supplementary Figure 6. HbF reactivation in SCD patient erythroid cells upon plasmid-mediated delivery of base editors to HSPCs.

a. RT-qPCR analysis of β^S- and γ-globin mRNA levels in SCD patient erythroblasts at day 13 of erythroid differentiation. β^S- and γ-globin mRNA expression was normalized to α-globin mRNA and expressed as percentage of the β^S-+γ- globins mRNA. The base-editing efficiency is indicated for each sample in the lower part of the panel. Data are expressed as single values or as mean ± SEM (n=2-3 biologically independent experiments, 2 donors). * p=0.0477; ** p=0.0080 for Donor 1, or p=0.0023 for Donor 2 LRF 4C, or p=0.0046 for Donor 2 LRF 8C; **** p≤0.0001 (Two-way ANOVA with Tukey correction for multiple comparisons).

b. Expression of γ- and β^S-globin chains measured by RP-HPLC in SCD patient erythroblasts at day 16 (donor 1) or day 19 (donor 2) of erythroid differentiation. β-like globin expression was normalized to α-globin. The base-editing efficiency is indicated for each sample in the lower part of the panel. Data are expressed as single values or as mean ± SEM (n=2-3 biologically independent experiments, 2 donors). ** p=0.0030 for Donor 1 LRF 4C, or p=0.0020 for Donor 1 LRF 8C, or p=0.0012 for Donor 2 LRF 4C, or p=0.0028 for Donor 2 LRF 8C; **** p≤0.0001 (Two-way ANOVA with Tukey correction for multiple comparisons).

Source data are provided as a Source Data file.

Supplementary Figure 7



Supplementary Figure 7. Analysis of BFU-E and CFU-GM colonies derived from plasmid-transfected SCD HSPCs.

a. Frequency of CFC in control and edited samples. SCD HSPCs mock-treated or treated with a sgRNA targeting the unrelated *AAVS1* locus served as controls. Data are expressed as mean ± SEM [BFU-E: n=9 (LRF 4C), n=3 (other groups) biologically independent experiments, 3 donors; CFU-GM: n=6 (LRF 4C), n=2 (other groups) biologically independent experiments, 3 donors]. No statistical differences were observed between control and edited samples (Two-way ANOVA, multiple comparisons).

b-c. Frequency of InDels in BFU-E (b) and CFU-GM (c) pooled colonies, measured by TIDE analysis, for edited and control (transfected with TE buffer or with CBE-SpRY plasmid and a sgRNA targeting the unrelated *AAVS1* locus) samples. The insertion or deletion of a C (±1 nt) in the homopoly-C stretch of the LRF 2T profile was separated from the overall frequency of InDels, as it was considered a sequencing error (**Supplementary Note 2**). Data are expressed as mean ± SEM [BFU-E: n=8 (LRF 4C), n=3 (other groups) biologically independent experiments, 3 donors; CFU-GM: n=6 (LRF 4C), n=2 (other groups) biologically independent experiments, 3 donors]. **** $p \leq 0.0001$ (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

d. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software in samples subjected to Sanger sequencing. Data are expressed as mean ± SEM [BFU-E: n=8 (LRF 4C), n=3 (other groups) biologically independent experiments, 3 donors; CFU-GM: n=6 (LRF 4C), n=2 (other groups) biologically independent experiments, 3 donors].

e. Expression of γ- and β^S-globin chains measured by RP-HPLC in BFU-E bulk populations. β-like globin expression was normalized to α-globin. Data are expressed as mean ± SEM [n=6 (LRF 4C), n=2 (other groups) biologically independent experiments, 2 donors]. ** p=0.0013; *** p=0.0002 for LRF 2T high, or p=0.0003 for KLF1 med, or p=0.0001 for KLF1 high (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

f. Analysis of HbF and HbS by cation-exchange HPLC in BFU-E bulk populations. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM [n=6 (LRF 4C), n=2 (other groups) biologically independent experiments, 2 donors]. ** p=0.0095; **** $p \leq 0.0001$ (Two-way ANOVA with Dunnett correction for multiple comparisons).

Source data are provided as a Source Data file.

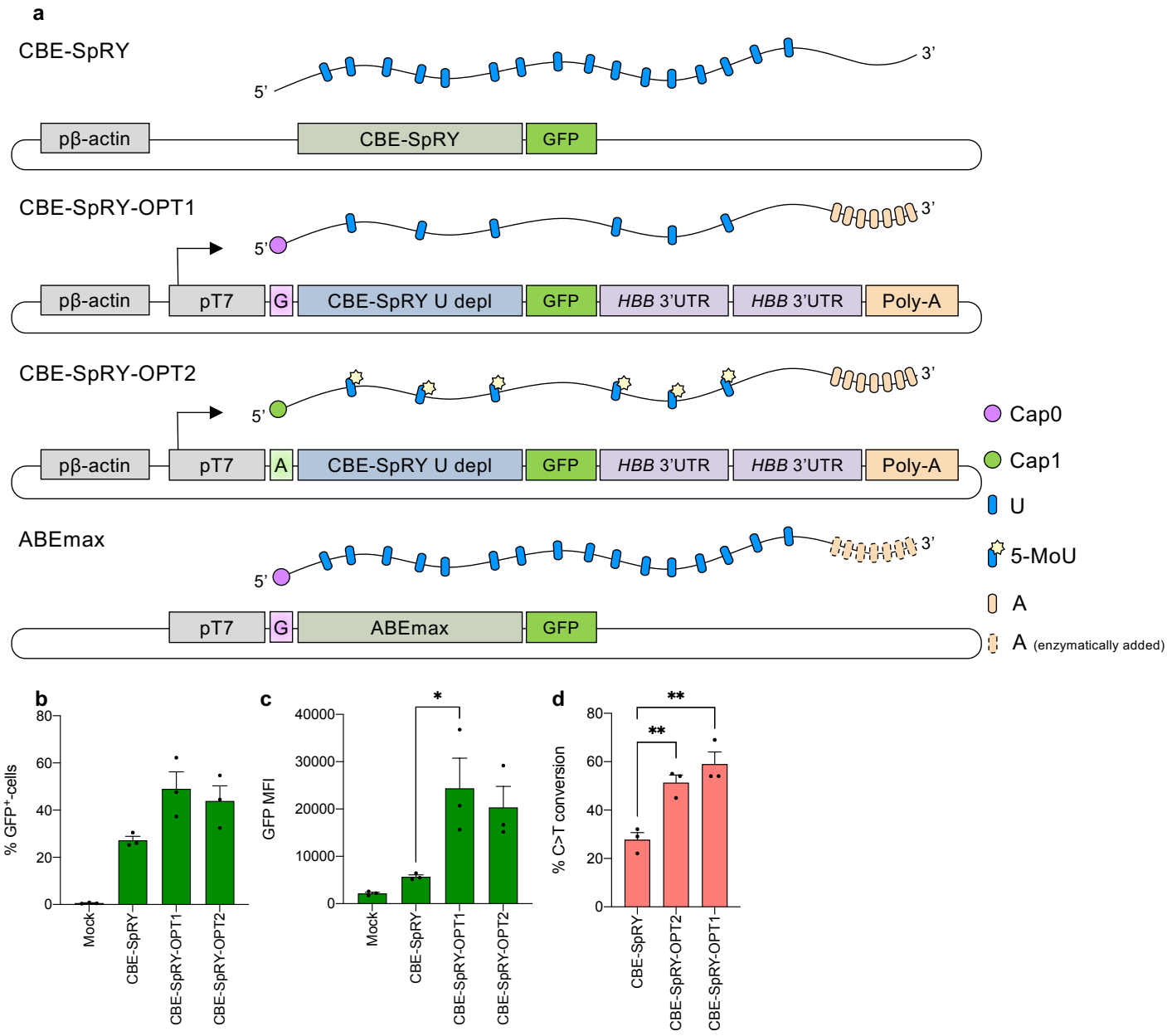
Supplementary Figure 8



Supplementary Figure 8. Targeted NGS sequencing of the *HBG1/2* promoters in erythroid BFU-E single colonies derived from SCD HSPCs transfected with ABE8e- and KLF_bs_1 sgRNA- expressing plasmids.

Frequency and sequence of modified and unmodified promoters in erythroid BFU-E single colonies (#101-119) derived from ABE8e-treated SCD HSPCs, as measured by targeted NGS sequencing. Data are expressed as single values (n=1 BFU-E colony). The experiment was performed once.

Supplementary Figure 9



Supplementary Figure 9. Plasmid optimization for CBE-SpRY *in vitro* transcription.

a. Schematic representation of CBE-SpRY plasmid optimization for *in vitro* transcription (ivt) and mRNA production. pβ-actin and pT7 are depicted by grey boxes. CBE-SpRY, ABEmax and GFP coding sequences, *HBB* 3'UTR and Poly-A are depicted by colored boxes. The mRNA produced upon ivt is represented above each plasmid construct. β-actin promoter (pβ-actin); T7 promoter (pT7); coding sequence of CBE-SpRY (CBE-SpRY); coding sequence of ABEmax (ABEmax); coding sequence of GFP (GFP); uridine depleted coding sequence of CBE-SpRY (CBE-SpRY U depl); *HBB* 3' untranslated region (*HBB* 3'UTR); Polyadenine tail (Poly-A); Guanine (G); Adenine (A); Uridine (U); 5-methoxyuridine (5-MoU).

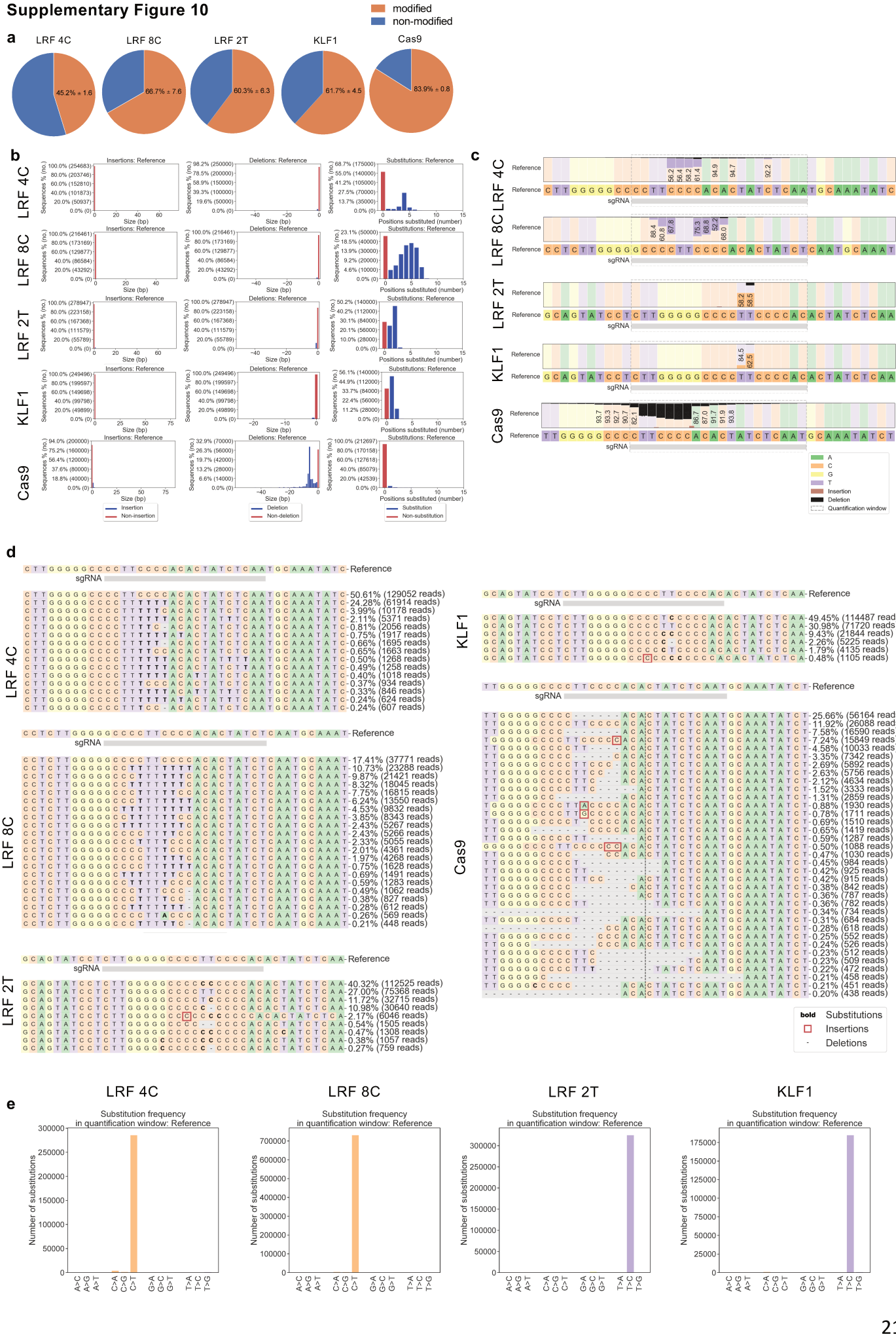
b. Frequency of GFP⁺ K562 cells upon transfection with different constructs depicted in A. We combined CBEs with LRF_bs_2 sgRNA. Data are expressed as mean ± SEM (n=3 biologically independent experiments). No statistical differences were observed between CBE-SpRY and CBE-SpRY-OPT1 and CBE-SpRY-OPT2 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

c. GFP MFI of GFP⁺ K562 cells upon transfection with different constructs depicted in A. Data are expressed as mean ± SEM (n=3 biologically independent experiments). * p=0.0452 (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).

d. C-G to T-A base-editing efficiency, calculated by the EditR software in samples subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=3 biologically independent

experiments). ** p=0.0021 for CBE-SpRY-OPT1, or p=0.0084 for CBE-SpRY-OPT2 (Ordinary One-way ANOVA).
Source data are provided as a Source Data file.

Supplementary Figure 10



Supplementary Figure 10. Targeted NGS sequencing of the *HBG1/2* promoters in erythroblasts derived from SCD HSPCs transfected with BE mRNA and chemically modified sgRNAs.

a. Frequency of modified and unmodified promoters (reads) in base editor- and Cas9-treated samples, as measured by targeted NGS sequencing. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).

b. Representative graphs (n=3 biologically independent experiments, 3 donors) showing the frequency of insertions, deletions and substitutions in base editor- and Cas9-treated samples, as measured by targeted NGS sequencing.

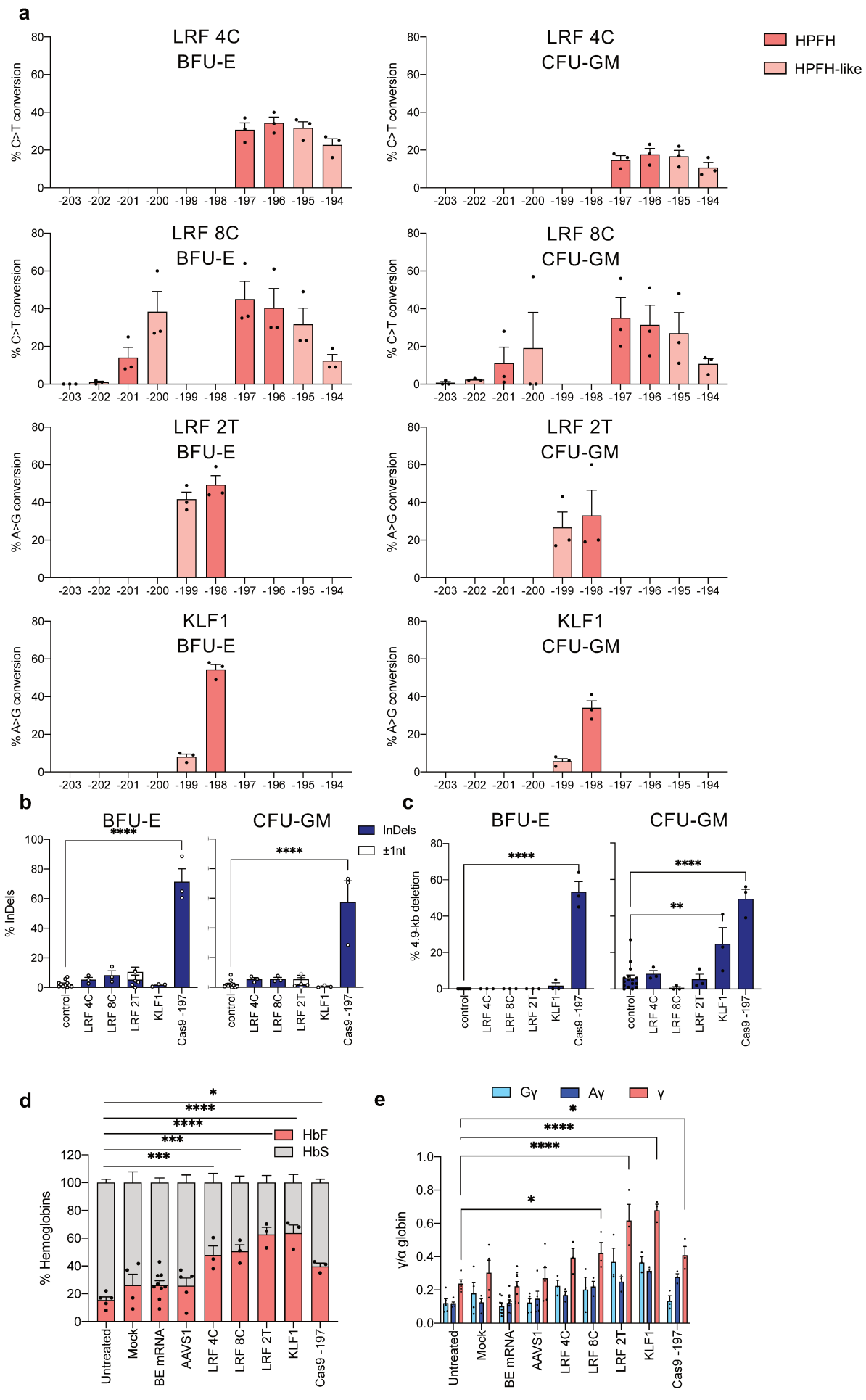
c. Representative graphs (n=3 biologically independent experiments, 3 donors) showing the frequency and location of insertions, deletions and substitutions in base editor- and Cas9-treated samples, as measured by targeted NGS sequencing.

d. Representative graphs (n=3 biologically independent experiments, 3 donors) showing the frequency and sequence of modified and unmodified promoters in base editor- and Cas9-treated samples, as measured by targeted NGS sequencing.

e. Representative graphs (n=3 biologically independent experiments, 3 donors) showing the product purity of base-editing enzymes, as indicated by the type of substitution and measured by targeted NGS sequencing.

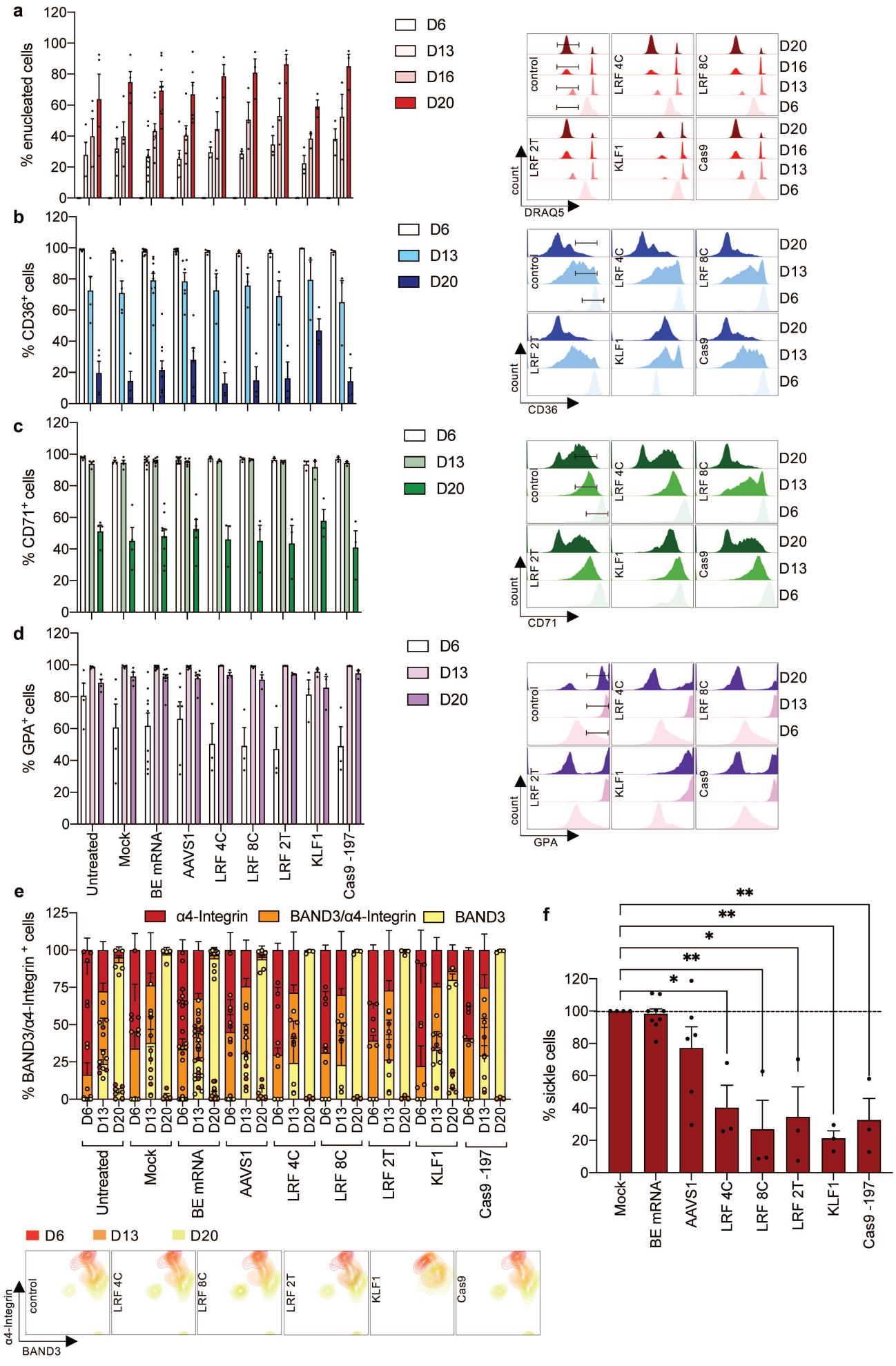
Source data are provided as a Source Data file.

Supplementary Figure 11



Supplementary Figure 11. CFC assay from RNA-transfected SCD HSPCs
a. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software in BFU-E and CFU-GM pooled colonies subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors).
b. Frequency of InDels in BFU-E and CFU-GM pooled colonies, measured by TIDE analysis, for edited and control (transfected with TE buffer, or with a BE-expressing plasmid alone or with CBE-SpRY plasmid and a sgRNA targeting the unrelated *AAVS1* locus) samples subjected to Sanger sequencing. The insertion or deletion of a C (±1 nt) in the homopoly-C stretch of the LRF 2T profile was separated from the overall frequency of InDels, as it was considered a sequencing error (**Supplementary Note 2**). Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). **** $p \leq 0.0001$ (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
c. Frequency of the 4.9-kb deletion in BFU-E and CFU-GM pooled colonies, measured by ddPCR, for edited and control (transfected with TE buffer, or with a base editor enzyme plasmid alone or with CBE-SpRY plasmid and a sgRNA targeting the unrelated *AAVS1* locus) samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). ** p=0.0030; **** $p \leq 0.0001$ (Ordinary One-way ANOVA with Dunnett correction for multiple comparisons).
d. Analysis of HbF and HbS by cation-exchange HPLC in BFU-E. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SD (n=3 biologically independent experiments, 3 donors). * p=0.0155; *** p=0.0005 for LRF 4C, or p=0.0001 for LRF 8C; **** $p \leq 0.0001$ (Two-way ANOVA with Sidak correction for multiple comparisons).
e. Expression of $^{G}\gamma$-, $^{A}\gamma$-, γ- ($^{G}\gamma$- + $^{A}\gamma$-) and β^{S}-globin chains measured by RP-HPLC in BFU-E. β-like globin expression was normalized to α-globin. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). * p=0.0281 for LRF 8C, or p=0.0468 for Cas9 -197; **** $p \leq 0.0001$ (Two-way ANOVA with Dunnett correction for multiple comparisons).
Source data are provided as a Source Data file.

Supplementary Figure 12



Supplementary Figure 12. Erythroid differentiation of SCD HSPCs upon RNA-mediated delivery of base editors.

a. Frequency of enucleated cells at day 6, 13, 16 and 20 of erythroid differentiation, as measured by flow cytometry analysis of DRAQ5 nuclear staining in control (untreated, or transfected with TE buffer, or transfected with a BE mRNA only, or transfected with a BE mRNA and a sgRNA targeting the unrelated *AAVS1* locus) and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). Representative flow cytometry histograms showing the $DRAQ5^-$ cell population for control and edited samples are reported.

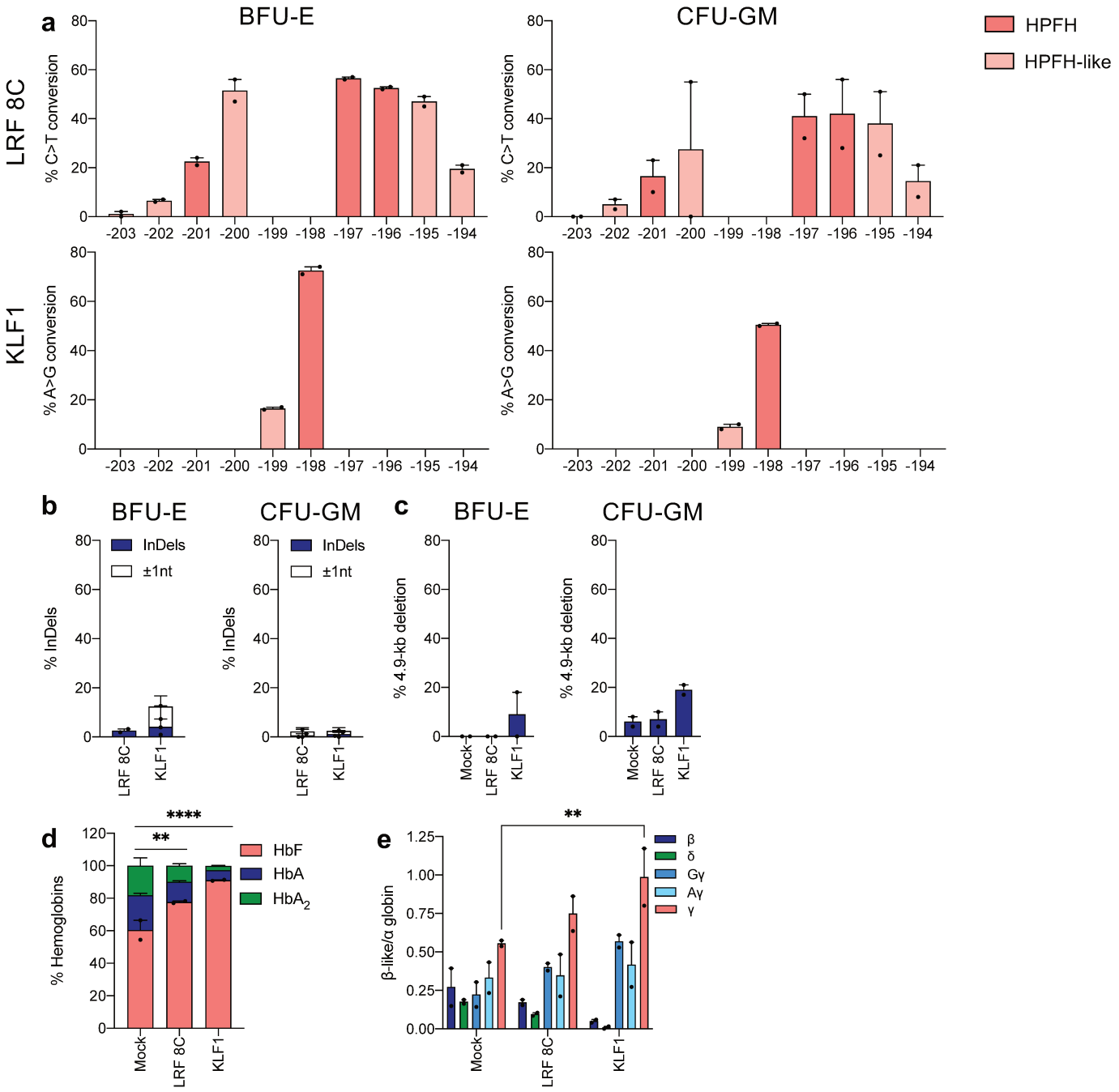
b-d. Frequency of $CD36^+$ (b), $CD71^+$ (c) and $GYPA^+$ (d) cells at day 6, 13 and 20 of erythroid differentiation, as measured by flow cytometry analysis of CD36, CD71 and GYPA erythroid markers. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). Representative flow cytometry histograms showing the $CD36^+$ (b), $CD71^+$ (c) and $GYPA^+$ (d) cell population for control and edited samples are reported.

e. Frequency of α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ in $7AAD^-$/$GYPA^+$ cells at day 6, 13 and 20 of erythroid differentiation, as measured by flow cytometry analysis of α4-Integrin and BAND3 erythroid markers. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). Representative flow cytometry contour plots showing the α4-$Integrin^+$, $BAND3^+$ and α4-$Integrin^+$/$BAND3^+$ cell population for control and edited samples are reported.

f. Frequency of sickling cells upon O_2 deprivation in control and edited samples. Data are expressed as mean ± SEM (n=3 biologically independent experiments, 3 donors). * p=0.0245 for LRF 4C, or p=0.0107 for LRF 2T; ** p=0.0034 for LRF 8C, or p=0.0014 for KLF1, or p=0.0080 for Cas9 -197 (One-way ANOVA with Tukey correction for multiple comparisons).

Source data are provided as a Source Data file.

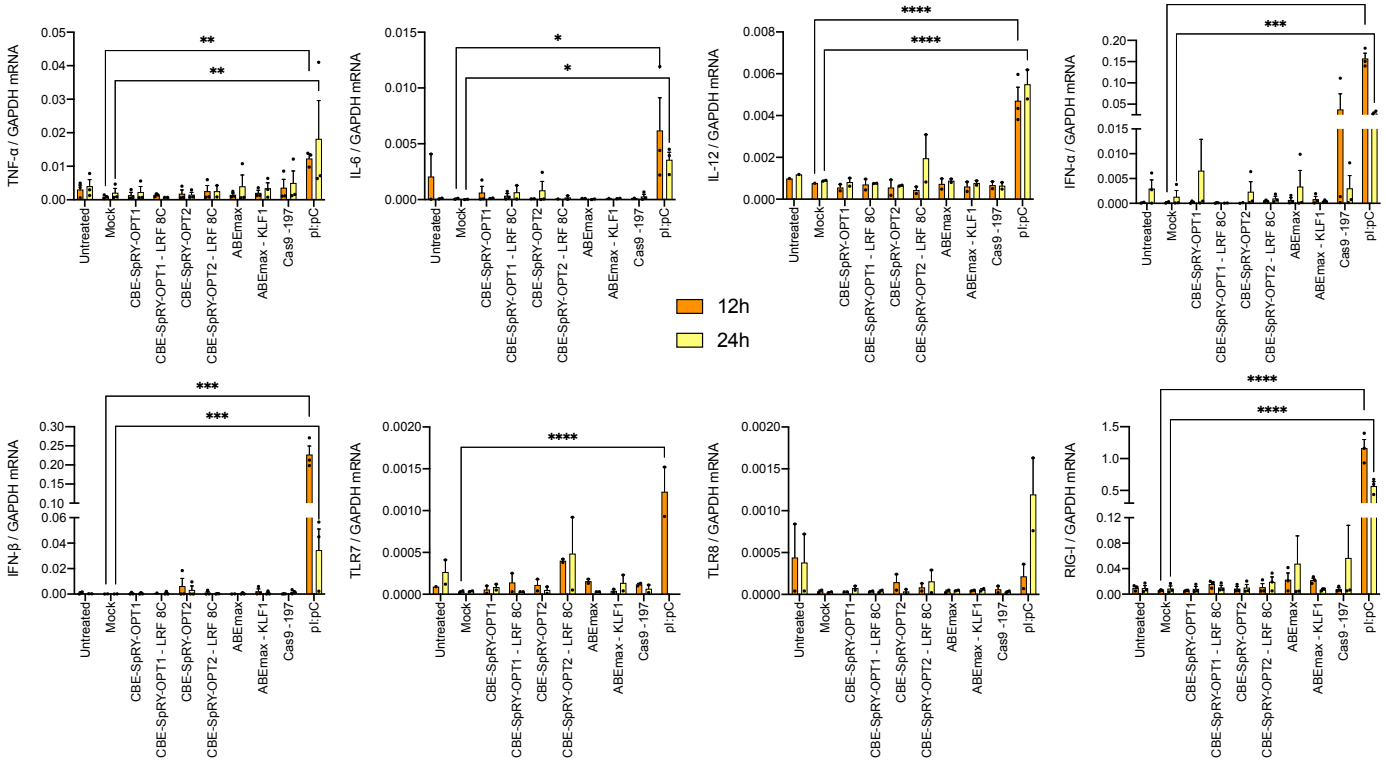
Supplementary Figure 13



Supplementary Figure 13. CFC assay from RNA-transfected β-thalassemic HSPCs.
a. C-G to T-A or A-T to G-C base-editing efficiency, calculated by the EditR software in BFU-E and CFU-GM pooled colonies subjected to Sanger sequencing. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors).
b. Frequency of InDels in BFU-E and CFU-GM pooled colonies, measured by TIDE analysis, for edited samples subjected to Sanger sequencing. The insertion or deletion of a C (±1 nt) in the homopoly-C stretch of the LRF 2T profile was separated from the overall frequency of InDels, as it was considered a sequencing error (**Supplementary Note 2**). Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors).
c. Frequency of the 4.9-kb deletion in BFU-E and CFU-GM pooled colonies, measured by ddPCR, for edited and control (transfected with TE buffer) samples. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). No statistical differences were observed between control and edited samples (Ordinary One-way ANOVA).

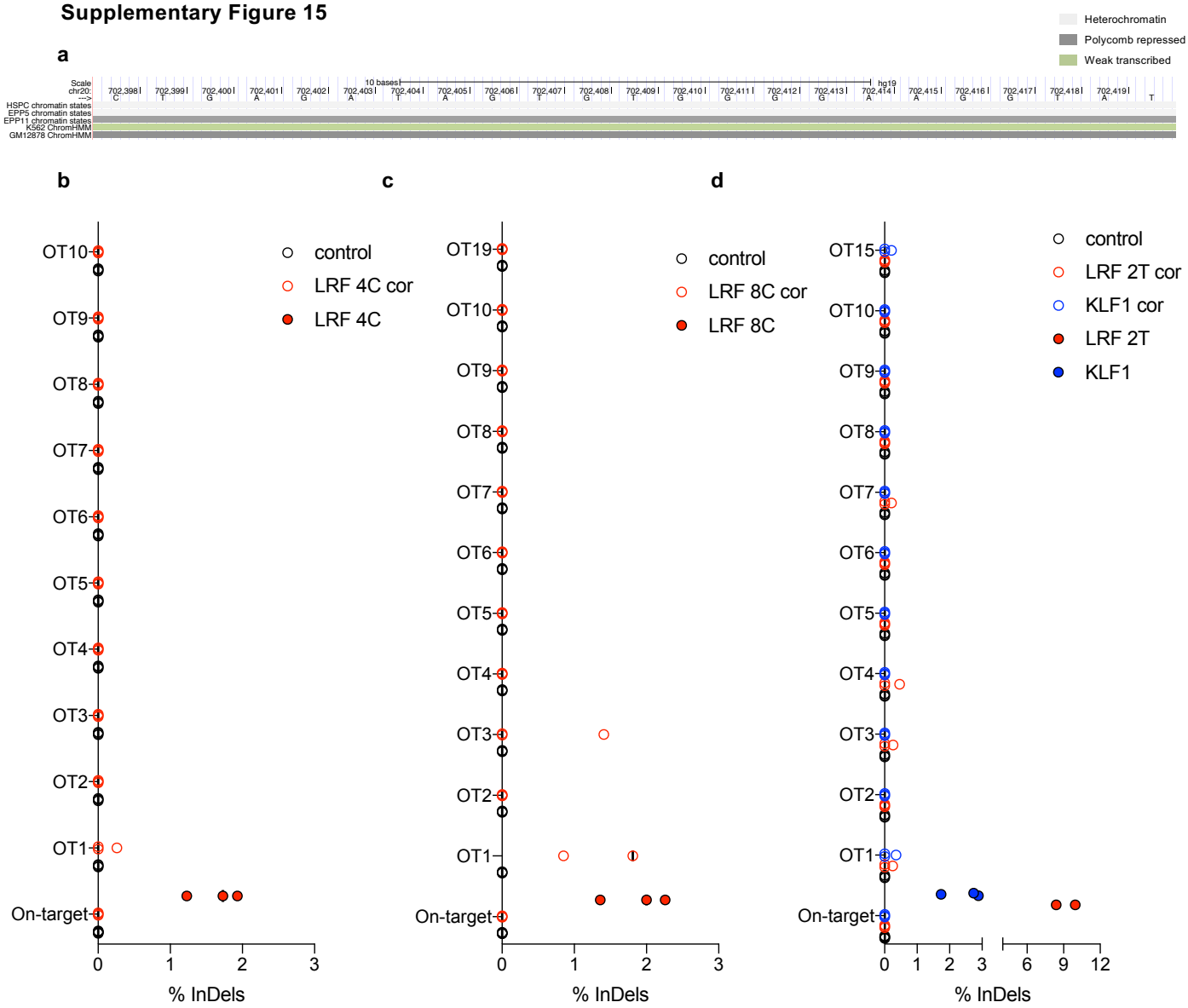
d. Analysis of HbF, HbA and HbA_2 by cation-exchange HPLC in BFU-E pooled colonies. We calculated the percentage of each Hb type over the total Hb tetramers. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). ** p=0.0026; **** p≤0.0001 (Two-way ANOVA with Dunnett correction for multiple comparisons).
e. Expression of β-, δ-, ^{G}γ-, ^{A}γ- and γ- (^{G}γ- + ^{A}γ-) globin chains measured by RP-HPLC in β-thalassemia patient RBCs. β-like globin expression was normalized to α-globin. The ratio α/non-α globins is reported on top of the graph. Data are expressed as mean ± SEM (n=2 biologically independent experiments, 2 donors). ** p=0.0078; (Two-way ANOVA with Dunnett correction for multiple comparisons).
Source data are provided as a Source Data file.

Supplementary Figure 14



Supplementary Figure 14. Analysis of genes activated by RNA stimuli.
RT-qPCR analysis performed 12 and 24 hours post-transfection in HD HSPCs. *TNF-α*, *IL-6*, *IL-12*, *IFN-α*, *IFN-β*, *TLR7*, *TLR8*, *RIG-I* mRNA expression was normalized to *GAPDH* mRNA. Polyinosinic:polycytidylic acid (pI:pC)-treated HD HSPCs were used as positive control. Data are expressed as mean ± SEM [*TNF-α*: n=2 (24h/CBE-SpRY-OPT1-LRF 8C and 24h/CBE-SpRY-OPT2-LRF 8C), n=3 (other groups) biologically independent experiments, 2-3 donors; *IL-6*: n=1 (12h/ABEmax-KLF1, 12h/CBE-SpRY-OPT2-LRF 8C and Ca9 -197), n=2 (Untreated, mock, 24h/CBE-SpRY-OPT1, 24h/CBE-SpRY-OPT1-LRF 8C, 12h/CBE-SpRY-OPT2, ABEmax, 24h/ABEmax-KLF1), n=3 (other groups) biologically independent experiments, 1-3 donors; *IL-12*: n=1 (Untreated and 12h/mock), n=3 (12h/pI:pC), n=2 (other groups) biologically independent experiments, 1-3 donors; *IFN-α*: n=2 (12h/Untreated, 12h/mock, 24h/CBE-SpRY-OPT1-LRF 8C and 12h/CBE-SpRY-OPT2), n=3 (other groups) biologically independent experiments, 2-3 donors; *IFN-β*: n=1 (12h/CBE-SpRY-OPT1-LRF 8C), n=2 (Untreated, mock and 12h/ABEmax), n=3 (other groups) biologically independent experiments, 1-3 donors; *TLR7*: n=1 (12h/Untreated), n=2 (other groups) biologically independent experiments, 1-2 donors; *TLR8*: n=2 biologically independent experiments, 2 donors; *RIG-I*: n=3 biologically independent experiments, 3 donors]. * p=0.0104; ** p=0.0027; *** p=0.0003 for *IFN-α*, or p=0.0002 for *IFN-β*; **** p≤0.0001 (Two-way ANOVA with Sidak correction for multiple comparisons).
Source data are provided as a Source Data file.

Supplementary Figure 15



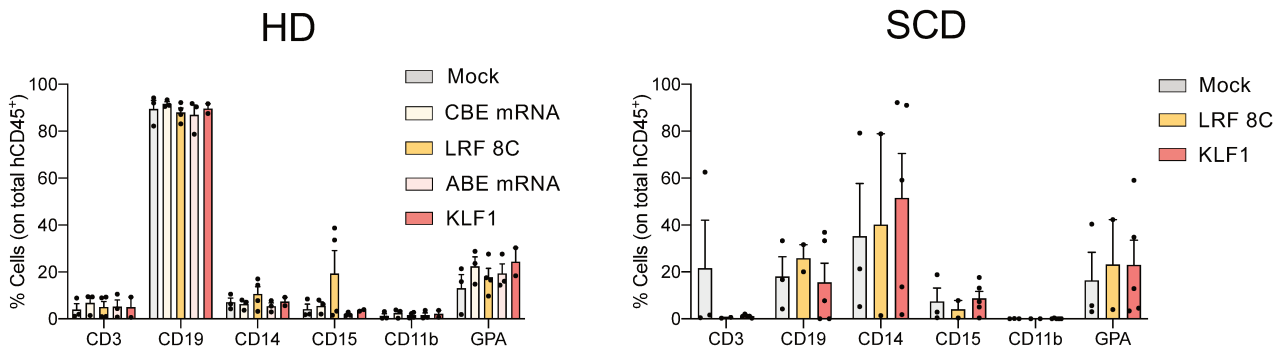
Supplementary Figure 15. sgRNA-dependent DNA off-target cleavage activity of CBE or ABE targeting the HBG promoters.

a. Analysis of the chromatin states at the LRF 8C-OT1 sequence in primary human HSPCs and HSPC-derived early (EPP5) and late (EPP11) erythroid precursors, and in erythroid (K562) and granulo-monocytic (GM12878) cell lines (UCSC datasets).

b-d. Frequency of InDels at on-target and off-target (OT) sites, for control and LRF 4C (b), LRF 8C (c), LRF 2T (d), and KLF1 (d) samples, as measured by targeted NGS sequencing. The frequency of InDels for the on-target site was corrected (cor) upon subtraction of the frequency of 1 nt insertion or deletion, that is due to sequencing error because of the presence of homopolymer C or T stretches. Data are expressed as individual values and median (n=3 biologically independent experiments, 3 donors).

Source data are provided as a Source Data file.

Supplementary Figure 16



Supplementary Figure 16. Human hematopoietic cell reconstitution in NBSGW mice transplanted with control and edited HSPCs.

Frequency of human T (CD3) and B (CD19) lymphoid, myeloid (CD14, CD15 and CD11b) and erythroid (GYPA) cells in BM in mice transplanted with control and edited HSPCs [HD: n=3 (Mock, CBE mRNA and ABE mRNA), n=4 (LRF 8C), n=2 (KLF1); SCD: n=3 (Mock), n=2 (LRF 8C), n=5 (KLF1) mice per group]. Each data point represents an individual mouse. Data are expressed as mean ± SEM.

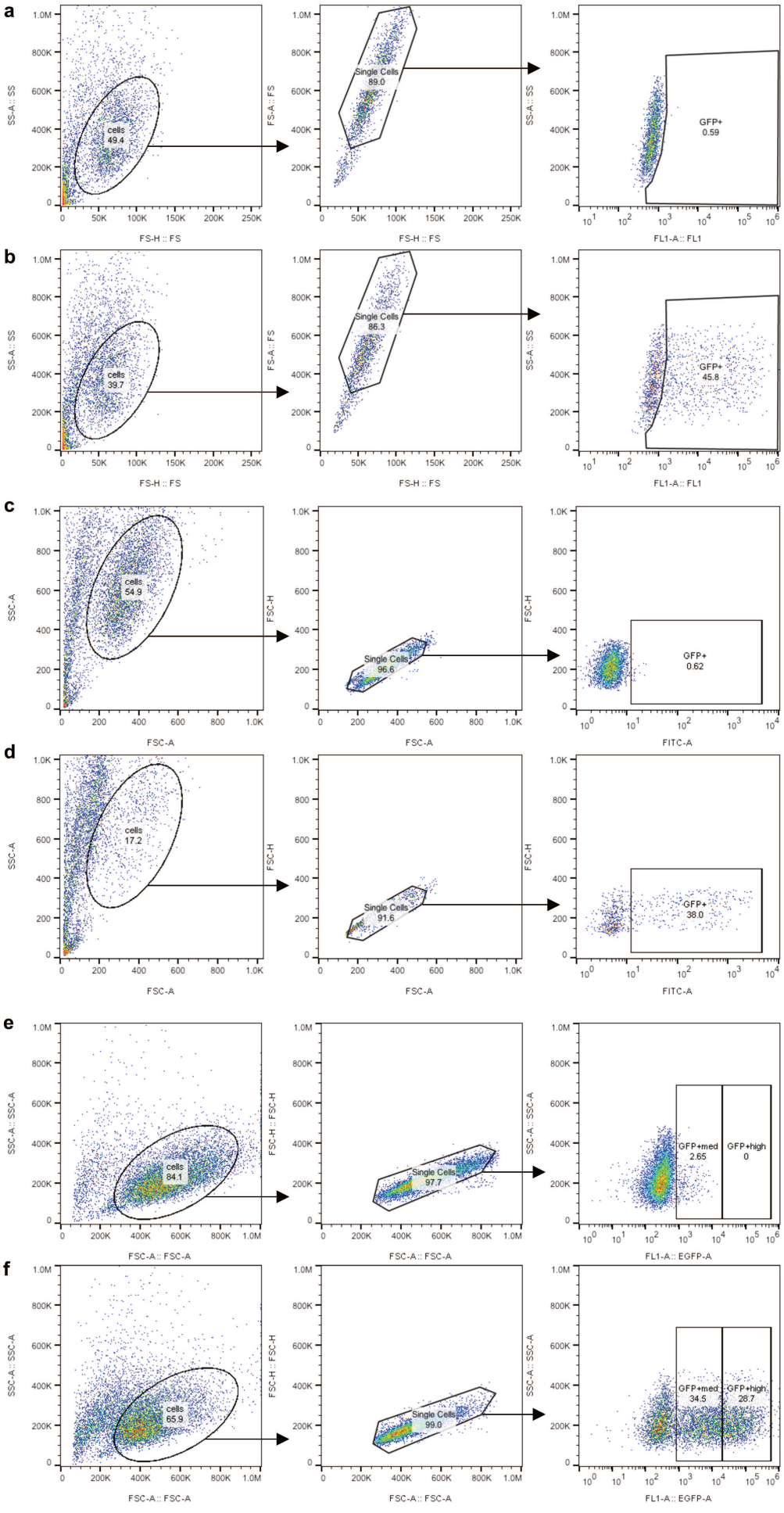
Source data are provided as a Source Data file.

Supplementary Figure 17



Supplementary Figure 17. Detection of the 4.9-kb deletion in repopulating HSCs.
PCR analysis of the 4.9-kb deletion in input and bone marrow-derived human HD and SCD cells. We used primers amplifying a 1435 bp-long product (del) only when the 4.9-kb deletion is present. Control primers amplifying the human *HBG* promoters (HBG1+HBG2 promoters) were used as DNA loading control (384 bp-long amplicon, h ctr). The different input and bone marrow (BM) HD and SCD cell samples are defined on top of each gel picture. This experiment was performed once.

Supplementary Figure 18

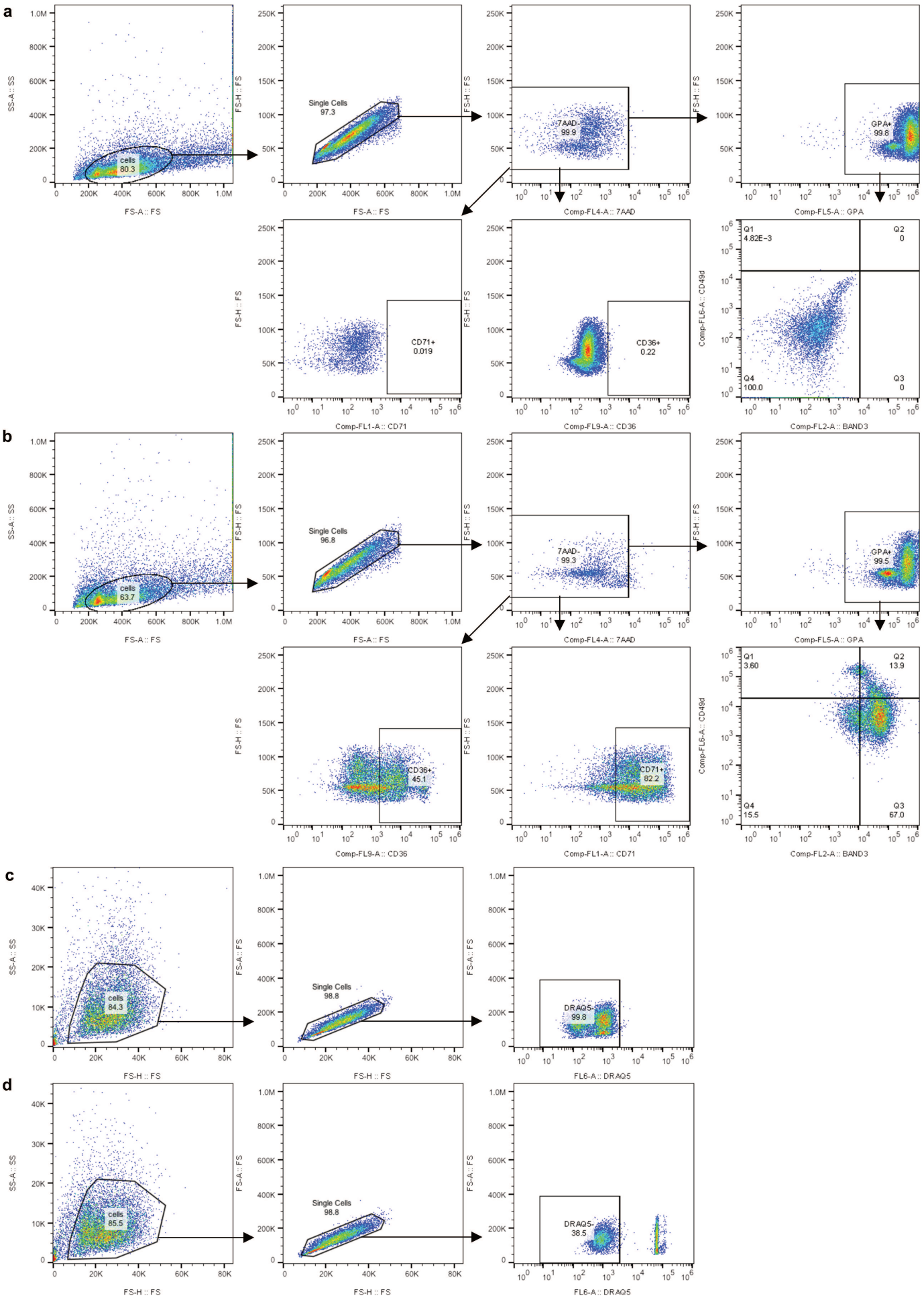


Supplementary Figure 18. Examples of gates used to assess transfection efficiency or to flow sort GFP$^+$ cells.

a-b. Representative gating strategy for population analysis on live, single, K562 cells in order to determine GFP expression via flow cytometry, in mock-transfected (**a**) and cells transfected with plasmids encoding fusions of base editors and GFP (**b**). This gating strategy was used to analyze data shown in Supplementary Figure 9b-c.

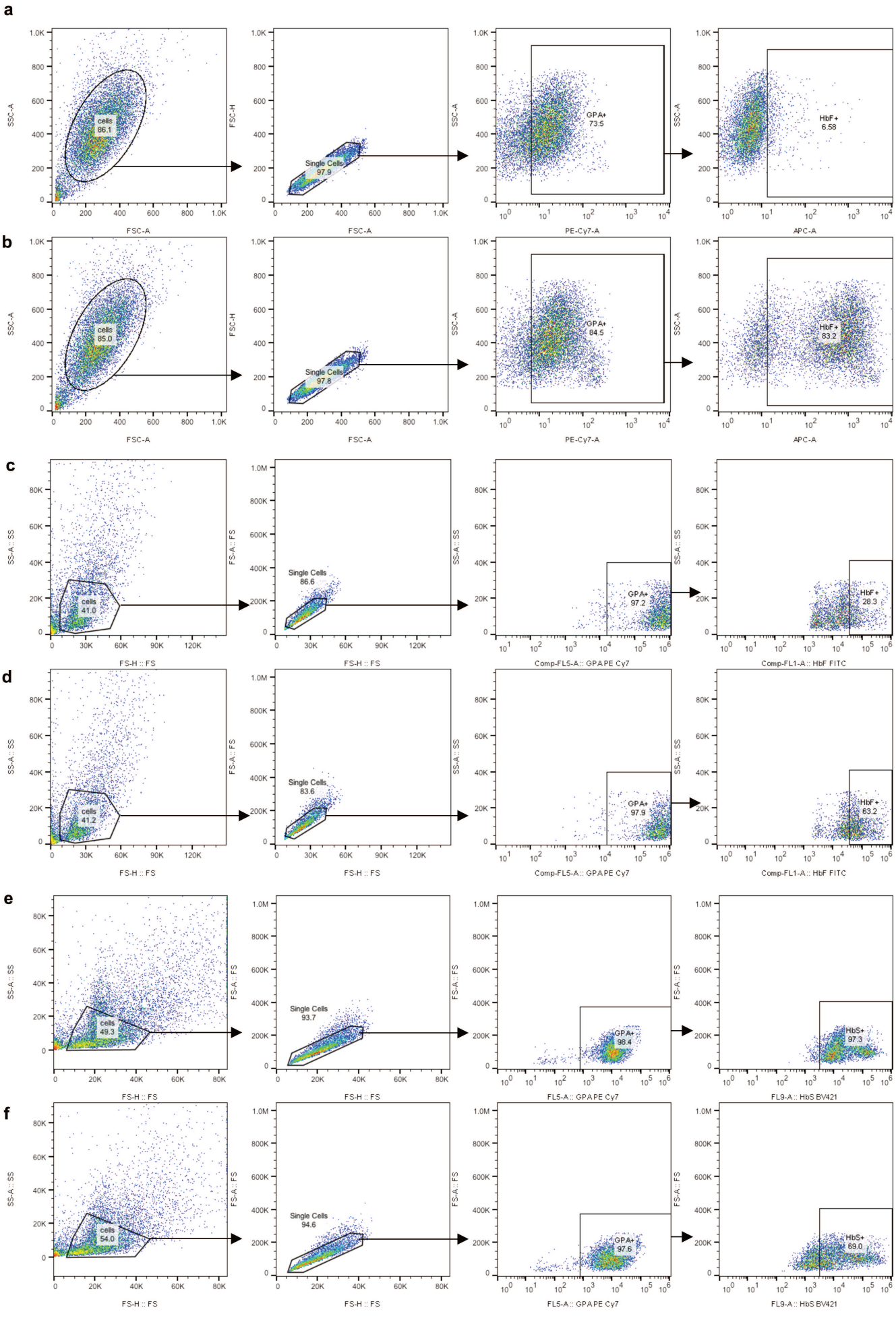
c-f. Representative gating strategy for population flow sorting of live, single, GFP$^+$ HUDEP2 cells (**c-d**), or live, single, GFP$^+$ HSPCs (**e-f**), in mock-transfected (**c and e**) and cells transfected with plasmids encoding fusions of base editors and GFP (**d and f**). This gating strategy was used to sort cell populations that were further analyzed for the production of data shown in Supplementary Figure 2 (**c-d**), or Figure 2 and Supplementary Figures 3-8 (**e-f**).

Supplementary Figure 19



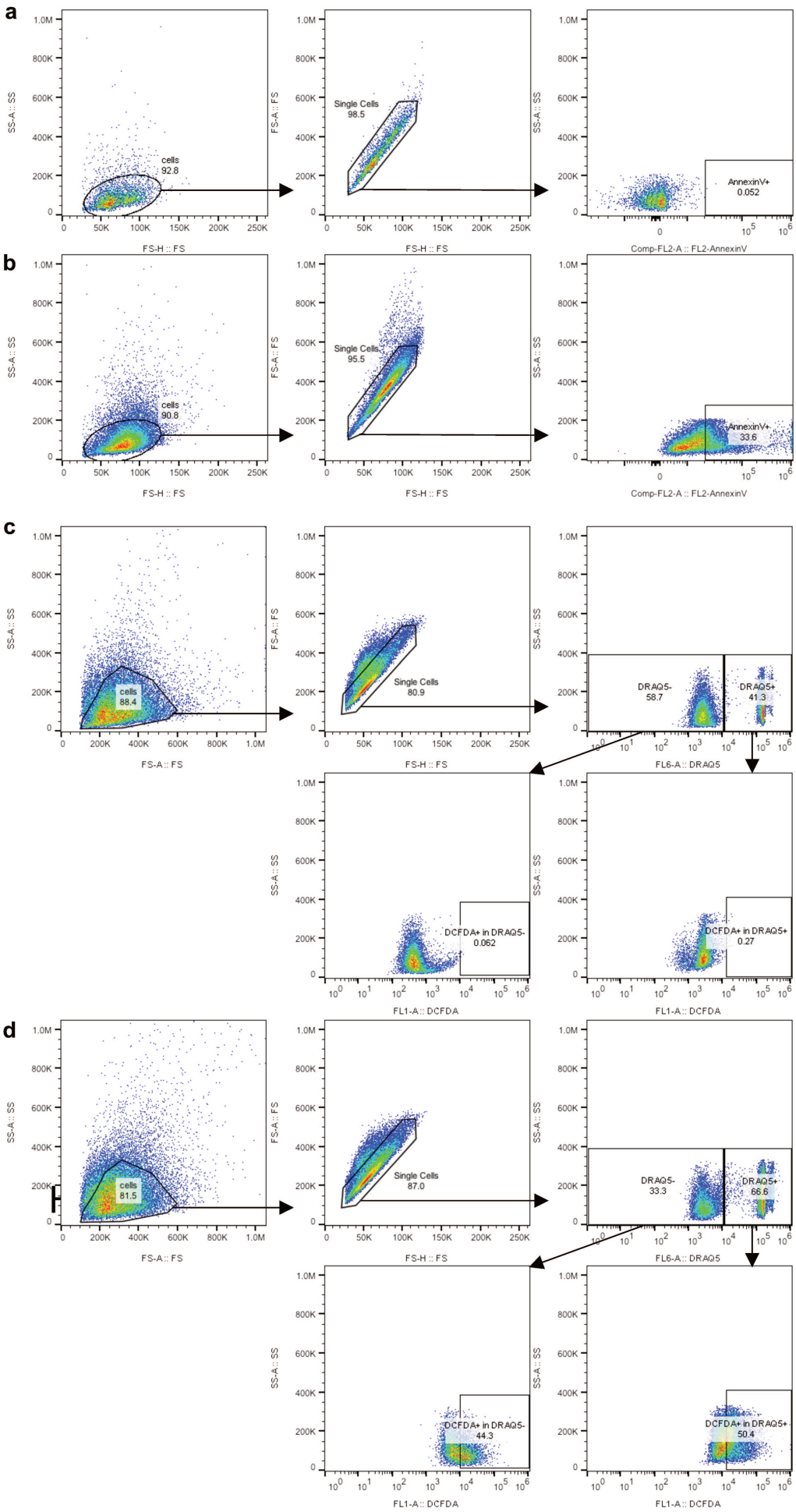
Supplementary Figure 19. Examples of gates used to assess erythroid surface markers and enucleated cells in HSPC-derived erythroid cells.
a-b. Representative gating strategy for population analysis on single, live ($7AAD^-$), erythroid cells in order to determine GPA, CD36, CD71, BAND3 and α4-Integrin expression via flow cytometry, in GPA-mono-stained (**a**) and stained cells (**b**). This gating strategy was used to analyze data shown in Figure 4l-o and Supplementary Figures 5b-e and 12b-e.
c-d. Representative gating strategy for population analysis on live, single, erythroid cells in order to determine enucleated cells ($DRAQ5^-$) via flow cytometry, in unstained (**c**), and stained cells (**d**). This gating strategy was used to analyze data shown in Figure 4j and Supplementary Figures 5a and 12a.

Supplementary Figure 20



Supplementary Figure 20. Examples of gates used to assess HbF and HbS expression.
a-b. Representative gating strategy for population analysis on live, single, HUDEP2 cells in order to determine HbF expression via flow cytometry, in control (**a**) and edited cells (**b**). This gating strategy was used to analyze data shown in Supplementary Figure 2b-c.
c-f. Representative gating strategy for population analysis on live, single, HSPC-derived erythroid cells in order to determine HbF (**c and d**) or HbS (**e and f**) expression via flow cytometry, in control (**c and e**) and edited cells (**d and f**). This gating strategy was used to analyze data shown in Figures 2f, 3l-m and 4g.

Supplementary Figure 21

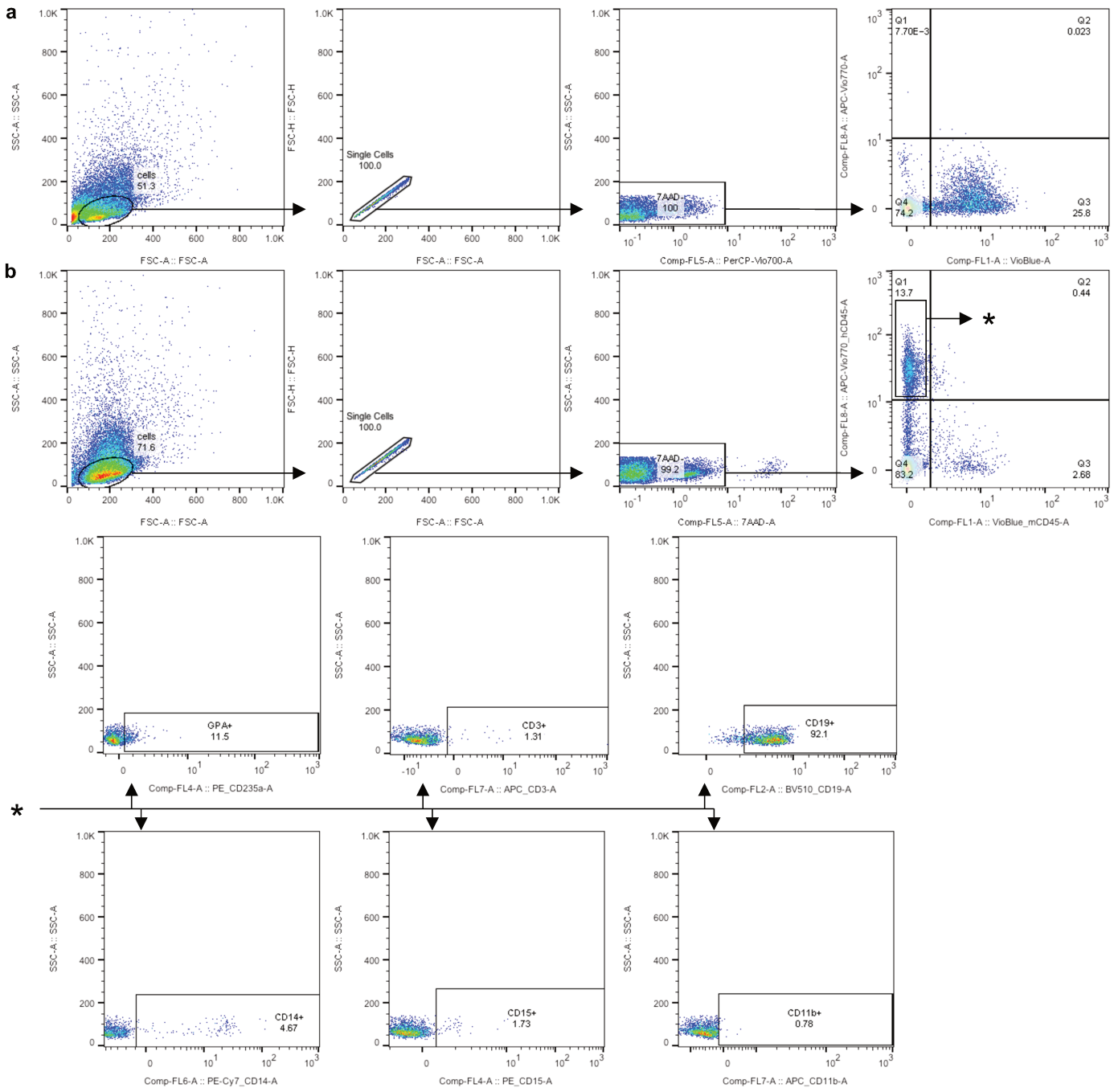


Supplementary Figure 21. Examples of gates used to assess apoptosis and ROS in HSPC-derived erythroid cells.

a-b. Representative gating strategy for population analysis on live, single, HSPC-derived erythroid cells in order to determine apoptotic cells via flow cytometry, in unstained (**a**) and stained cells (**b**). This gating strategy was used to analyze data shown in Figure 4p.

c-d. Representative gating strategy for population analysis on live, single, HSPC-derived nucleated ($DRAQ5^+$) and enucleated ($DRAQ5^-$) erythroid cells in order to determine ROS via flow cytometry, in DRAQ5-mono-stained (**c**) and stained cells (**d**). This gating strategy was used to analyze data shown in Figure 4q.

Supplementary Figure 22



Supplementary Figure 22. Examples of gates used to assess chimerism and lineage specific markers in xenotransplantation experiments.
Representative gating strategy for population analysis on live, single, bone marrow-, spleen-, thymus-, and blood-derived human-mouse chimeric cells in order to determine human chimerism and lineage specific markers expression in xenotransplanted mice via flow cytometry, in mCD45-mono-stained (**a**) and stained cells (**b**). This gating strategy was used to analyze data shown in Figure 8b-c and Supplementary Figure 16.

Supplementary sequences

CBE-SpRY-OPT sequence

T7 promoter
nucleotide for suitable capping (G or A for CBE-SpRY-OPT1 and CBE-SpRY-OPT2 respectively)
CBE-SpRY Coding sequence, Uridine depleted
2 copies of the HBB 3'UTR
Poly-A tail

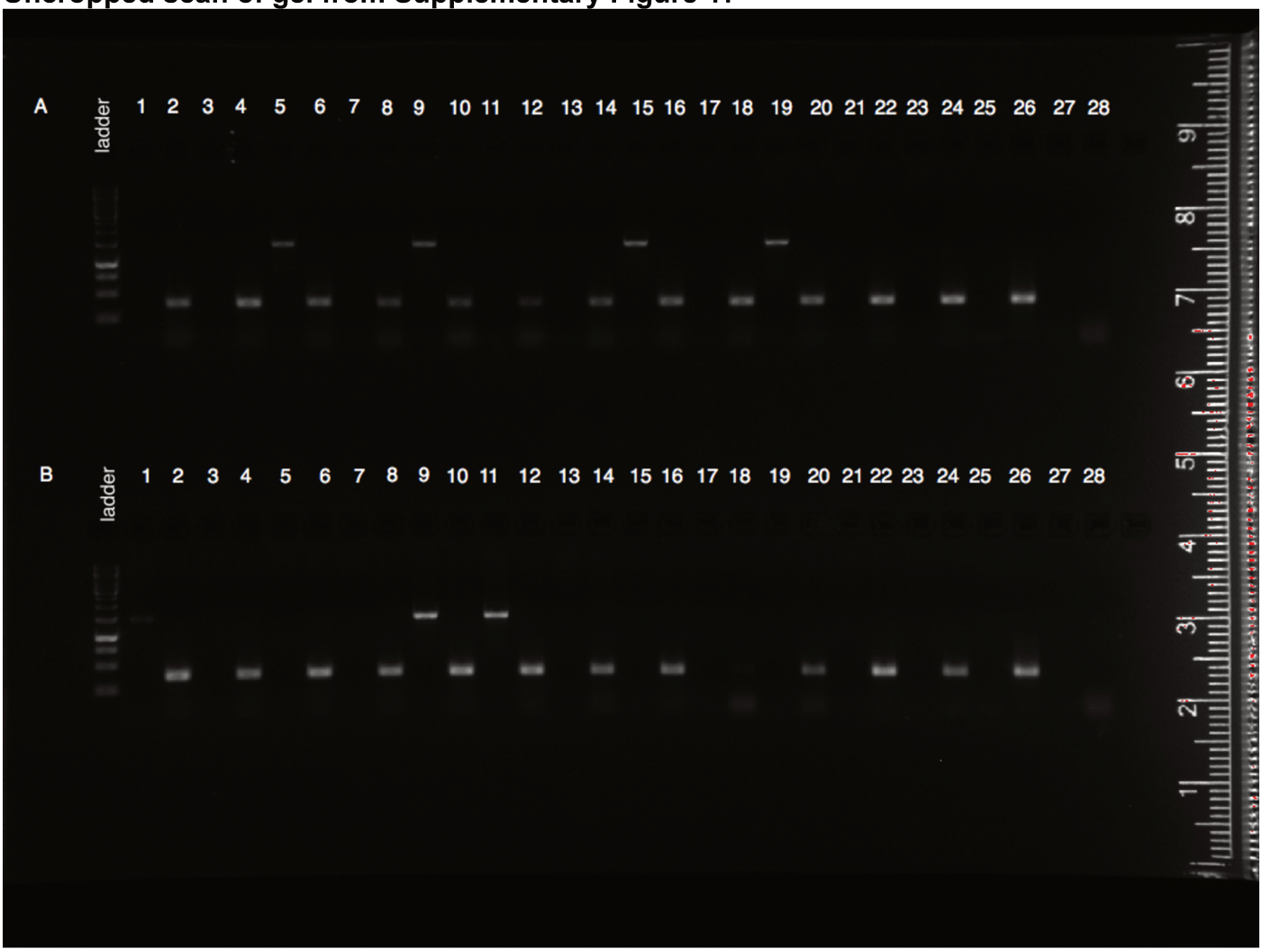
TCCGCGCACATTTCCCCGAAAAGTGCCACCTGGGTCGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAG
TTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGA
CGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCC
ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGT
TCTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGG
CGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAAT
CAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGA
GTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCA
CAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGT
GAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCC
GCGTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGC
GCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCA
GGGGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGG
CTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGG
GCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTA
TGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCT
AGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTT
CTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTG
ACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGTAATACGACTCACTATA[G/A]G
GAGAGCCGCCACCATGAAGCGGACCGCCGACGGCAGCGAGTTCGAGAGCCCCAAGAAGAAGCGGAAGGTGAGCAGCGAGACCGGCCC
CGTGGCCGTGGACCCCACCCTGCGGCGGCGGATCGAGCCCCACGAGTTCGAGGTGTTCTTCGACCCCCGGGAGCTGCGGAAGGAGAC
CTGCCTGCTGTACGAGATCAACTGGGGCGGCCGGCACAGCATCTGGCGGCACACCAGCCAGAACACCAACAAGCACGTGGAGGTGAA
CTTCATCGAGAAGTTCACCACCGAGCGGTACTTCTGCCCCAACACCCGGTGCAGCATCACCTGGTTCCTGAGCTGGAGCCCCTGCGG
CGAGTGCAGCCGGGCCATCACCGAGTTCCTGAGCCGGTACCCCCACGTGACCCTGTTCATCTACATCGCCCGGCTGTACCACCACGC
CGACCCCCGGAACCGGCAGGGCCTGCGGGACCTGATCAGCAGCGGCGTGACCATCCAGATCATGACCGAGCAGGAGAGCGGCTACTG
CTGGCGGAACTTCGTGAACTACAGCCCCAGCAACGAGGCCCACTGGCCCCGGTACCCCCACCTGTGGGTGCGGCTGTACGTGCTGGA
GCTGTACTGCATCATCCTGGGCCTGCCCCCCTGCCTGAACATCCTGCGGCGGAAGCAGCCCCAGCTGACCTTCTTCACCATCGCCCT
GCAGAGCTGCCACTACCAGCGGCTGCCCCCCCACATCCTGTGGGCCACCGGCCTGAAGAGCGGCGGCAGCAGCGGCGGCAGCAGCGG
CAGCGAGACCCCCGGCACCAGCGAGAGCGCCACCCCCGAGAGCAGCGGCGGCAGCAGCGGCGGCAGCGACAAGAAGTACAGCATCGG
CCTGGCCATCGGCACCAACAGCGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAGTTCAAGGTGCTGGGCAA
CACCGACCGGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTTCGACAGCGGCGAGACCGCCGAGCGGACCCGGCTGAAGCG
GACCGCCCGGCGGCGGTACACCCGGCGGAAGAACCGGATCTGCTACCTGCAGGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGA
CAGCTTCTTCCACCGGCTGGAGGAGAGCTTCCTGGTGGAGGAGGACAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGA
CGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGCGGAAGAAGCTGGTGGACAGCACCGACAAGGCCGACCTGCGGCT
GATCTACCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGA
CAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAGAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCAT
CCTGAGCGCCCGGCTGAGCAAGAGCCGGCGGCTGGAGAACCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAACGGCCTGTTCGGCAA
CCTGATCGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGACGCCAAGCTGCAGCTGAGCAAGGA
CACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTCCTGGCCGCCAAGAACCTGAGCGA
CGCCATCCTGCTGAGCGACATCCTGCGGGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCAGCATGATCAAGCGGTACGACGA
GCACCACCAGGACCTGACCCTGCTGAAGGCCCTGGTGCGGCAGCAGCTGCCCGAGAAGTACAAGGAGATCTTCTTCGACCAGAGCAA
GAACGGCTACGCCGGCTACATCGACGGCGGCGCCAGCCAGGAGGAGTTCTACAAGTTCATCAAGCCCATCCTGGAGAAGATGGACGG
CACCGAGGAGCTGCTGGTGAAGCTGAACCGGGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGAT
CCACCTGGGCGAGCTGCACGCCATCCTGCGGCGGCAGGAGGACTTCTACCCCTTCCTGAAGGACAACCGGGAGAAGATCGAGAAGAT
CCTGACCTTCCGGATCCCCTACTACGTGGGCCCCCTGGCCCGGGGCAACAGCCGGTTCGCCTGGATGACCCGGAAGAGCGAGGAGAC
CATCACCCCCTGGAACTTCGAGGAGGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGACAAGAA
CCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAGGTGAAGTACGT
GACCGAGGGCATGCGGAAGCCCGCCTTCCTGAGCGGCGAGCAGAAGAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAGGT
GACCGTGAAGCAGCTGAAGGAGGACTACTTCAAGAAGATCGAGTGCTTCGACAGCGTGGAGATCAGCGGCGTGGAGGACCGGTTCAA
CGCCAGCCTGGGCACCTACCACGACCTGCTGAAGATCATCAAGGACAAGGACTTCCTGGACAACGAGGAGAACGAGGACATCCTGGA
GGACATCGTGCTGACCCTGACCCTGTTCGAGGACCGGGAGATGATCGAGGAGCGGCTGAAGACCTACGCCCACCTGTTCGACGACAA
GGTGATGAAGCAGCTGAAGCGGCGGCGGTACACCGGCTGGGGCCGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGAG
CGGCAAGACCATCCTGGACTTCCTGAAGAGCGACGGCTTCGCCAACCGGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTT
CAAGGAGGACATCCAGAAGGCCCAGGTGAGCGGCCAGGGCGACAGCCTGCACGAGCACATCGCCAACCTGGCCGGCAGCCCCGCCAT
CAAGAAGGGCATCCTGCAGACCGTGAAGGTGGTGGACGAGCTGGTGAAGGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGA
GATGGCCCGGGAGAACCAGACCACCCAGAAGGGCCAGAAGAACAGCCGGGAGCGGATGAAGCGGATCGAGGAGGGCATCAAGGAGCT
GGGCAGCCAGATCCTGAAGGAGCACCCCGTGGAGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAACGGCCG

GGACATGTACGTGGACCAGGAGCTGGACATCAACCGGCTGAGCGACTACGACGTGGACCACATCGTGCCCCAGAGCTTCCTGAAGGA
CGACAGCATCGACAACAAGGTGCTGACCCGGAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCAGCGAGGAGGTGGTGAAGAA
GATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATCACCCAGCGGAAGTTCGACAACCTGACCAAGGCCGAGCGGGGCGG
CCTGAGCGAGCTGGACAAGGCCGGCTTCATCAAGCGGCAGCTGGTGGAGACCCGGCAGATCACCAAGCACGTGGCCCAGATCCTGGA
CAGCCGGATGAACACCAAGTACGACGAGAACGACAAGCTGATCCGGGAGGTGAAGGTGATCACCCTGAAGAGCAAGCTGGTGAGCGA
CTTCCGGAAGGACTTCCAGTTCTACAAGGTGCGGGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTGGTGGG
CACCGCCCTGATCAAGAAGTACCCCAAGCTGGAGAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGC
CAAGAGCGAGCAGGAGATCGGCAAGGCCACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTCTTCAAGACCGAGATCACCCT
GGCCAACGGCGAGATCCGGAAGCGGCCCCTGATCGAGACCAACGGCGAGACCGGCGAGATCGTGTGGGACAAGGGCCGGGACTTCGC
CACCGTGCGGAAGGTGCTGAGCATGCCCCAGGTGAACATCGTGAAGAAGACCGAGGTGCAGACCGGCGGCTTCAGCAAGGAGAGCAT
CCGGCCCAAGCGGAACAGCGACAAGCTGATCGCCCGGAAGAAGGACTGGGACCCCAAGAAGTACGGCGGCTTCCTGTGGCCCACCGT
GGCCTACAGCGTGCTGGTGGTGGCCAAGGTGGAGAAGGGCAAGAGCAAGAAGCTGAAGAGCGTGAAGGAGCTGCTGGGCATCACCAT
CATGGAGCGGAGCAGCTTCGAGAAGAACCCCATCGACTTCCTGGAGGCCAAGGGCTACAAGGAGGTGAAGAAGGACCTGATCATCAA
GCTGCCCAAGTACAGCCTGTTCGAGCTGGAGAACGGCCGGAAGCGGATGCTGGCCAGCGCCAAGCAGCTGCAGAAGGGCAACGAGCT
GGCCCTGCCCAGCAAGTACGTGAACTTCCTGTACCTGGCCAGCCACTACGAGAAGCTGAAGGGCAGCCCCGAGGACAACGAGCAGAA
GCAGCTGTTCGTGGAGCAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCAGCAAGCGGGTGATCCTGGCCGA
CGCCAACCTGGACAAGGTGCTGAGCGCCTACAACAAGCACCGGGACAAGCCCATCCGGGAGCAGGCCGAGAACATCATCCACCTGTT
CACCCTGACCCGGCTGGGCGCCCCCCGGGCCTTCAAGTACTTCGACACCACCATCGACCCCAAGCAGTACCGGAGCACCAAGGAGGT
GCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACCCGGATCGACCTGAGCCAGCTGGGCGGCGACAGCGGCGG
CAGCGGCGGCAGCGGCGGCAGCACCAACCTGAGCGACATCATCGAGAAGGAGACCGGCAAGCAGCTGGTGATCCAGGAGAGCATCCT
GATGCTGCCCGAGGAGGTGGAGGAGGTGATCGGCAACAAGCCCGAGAGCGACATCCTGGTGCACACCGCCTACGACGAGAGCACCGA
CGAGAACGTGATGCTGCTGACCAGCGACGCCCCCGAGTACAAGCCCTGGGCCCTGGTGATCCAGGACAGCAACGGCGAGAACAAGAT
CAAGATGCTGAGCGGCGGCAGCGGCGGCAGCGGCGGCAGCACCAACCTGAGCGACATCATCGAGAAGGAGACCGGCAAGCAGCTGGT
GATCCAGGAGAGCATCCTGATGCTGCCCGAGGAGGTGGAGGAGGTGATCGGCAACAAGCCCGAGAGCGACATCCTGGTGCACACCGC
CTACGACGAGAGCACCGACGAGAACGTGATGCTGCTGACCAGCGACGCCCCCGAGTACAAGCCCTGGGCCCTGGTGATCCAGGACAG
CAACGGCGAGAACAAGATCAAGATGCTGAGCGGCGGCAGCAAGCGGACCGCCGACGGCAGCGAGTTCGAGCCCAAGAAGAAGCGGAA
GGTGGGCGGCGGCGGCAGCGGCGCCACCAACTTCAGCCTGCTGAAGCAGGCCGGCGACGTGGAGGAGAACCCCGGCCCCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT
GGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGGTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCC
CGAGGGCTACGTGCAGGAGCGGACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGGGCCGAGGTGAAGTTCGAGGGCGACAC
CCTGGTGAACCGGATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAG
CCACAACGTGTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGGCACAACATCGAGGACGGCAGCGT
GCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAG
CGCCCTGAGCAAGGACCCCAACGAGAAGCGGGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGGGCATGGA
CGAGCTGTACAAGTGAGCTAGAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAA
CTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCTGCGCTAGAAGCTCGCTTT
CTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCT
GGATTCTGCCTAATAAAAAACATTTATTTTCATTGCTGCGGGACATTCTTAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCTTAAGTTAAAATAAGGCTAG
TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTAGAAGCGGCCGCACTCCTCAGGTGCAGGCTGCCTATCAGA
AGGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCATGA
AGCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCG
GAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTG
CCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGA
CTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTAGCCA
GATTTTTCCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCAAGCTTGGCG
TAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAA
GCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG
CGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATT
CTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAG
GCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
TTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCGCTG
CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCG
GTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAG
CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCA
CGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTT
TGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAA
AACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGT
TCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTA
TCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT

GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC
CCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGA
GAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTC
ATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGG
GCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC
ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGT

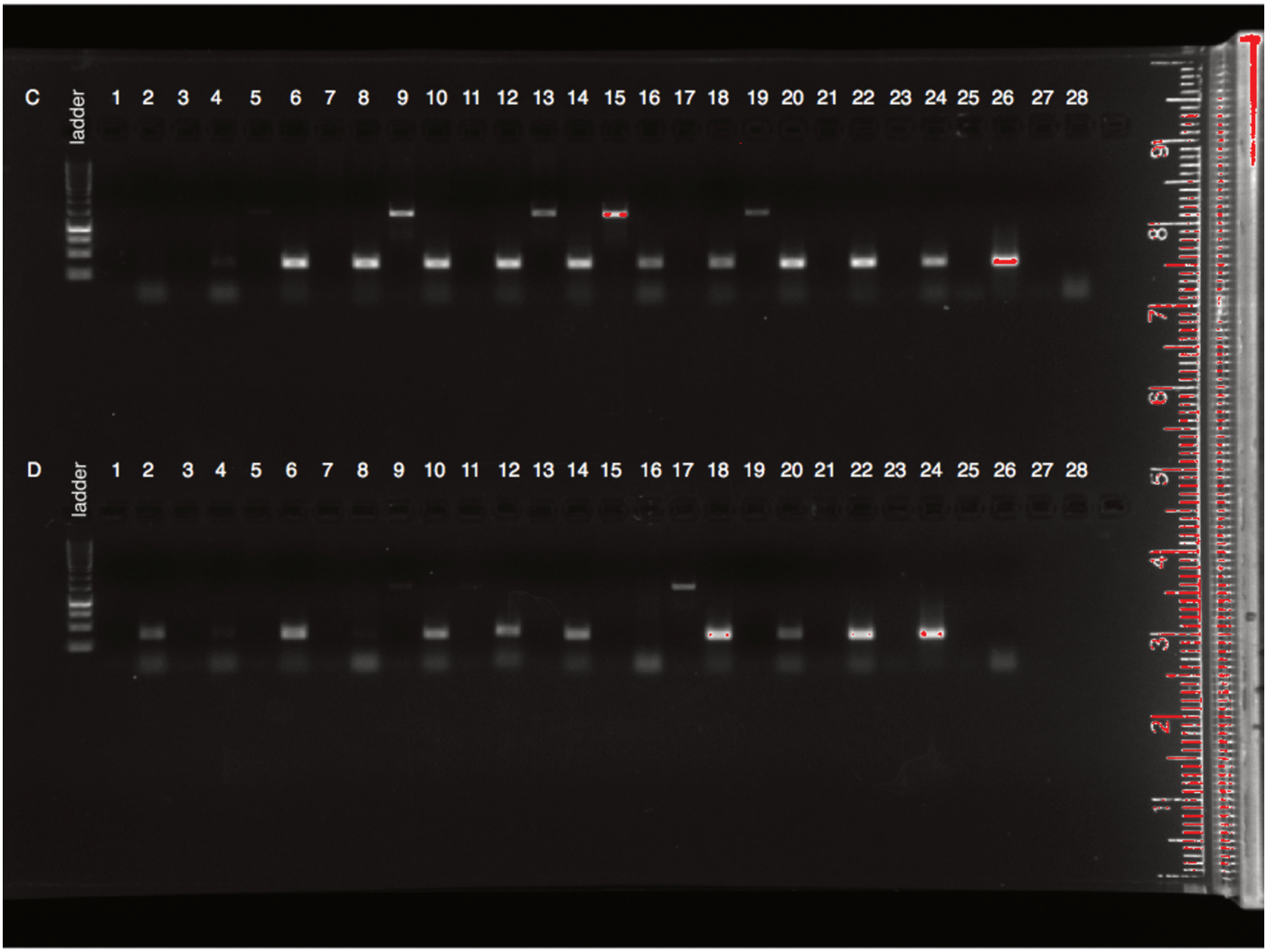
Supplementary uncropped scans of blot and gels

Uncropped scan of gel from Supplementary Figure 17



Row	Number	Sample				
A	1	HD	Input	Erythroid liquid culture	Mock	4.9-kb deletion
	2					human control
	3				CBE	4.9-kb deletion
	4					human control
	5				LRF 8C	4.9-kb deletion
	6					human control
	7				ABE	4.9-kb deletion
	8					human control
	9				KLF1	4.9-kb deletion
	10					human control
	11			HSPC medium	Mock	4.9-kb deletion
	12					human control
	13				CBE	4.9-kb deletion
	14					human control
	15				LRF 8C	4.9-kb deletion
	16					human control
	17				ABE	4.9-kb deletion
	18					human control
	19				KLF1	4.9-kb deletion
	20					human control
	21		Bone marrow	Mock		4.9-kb deletion
	22					human control
	23			CBE		4.9-kb deletion
	24					human control

	25	Unrelated samples			
	26				
	27	H_2O			4.9-kb deletion
	28				human control
B	1	HD	Bone marrow	LRF 8C	4.9-kb deletion
	2				human control
	3			ABE	4.9-kb deletion
	4				human control
	5			Mock	4.9-kb deletion
	6				human control
	7			CBE	4.9-kb deletion
	8				human control
	9			LRF 8C	4.9-kb deletion
	10				human control
	11			KLF1	4.9-kb deletion
	12				human control
	13			ABE	4.9-kb deletion
	14				human control
	15			Mock	4.9-kb deletion
	16				human control
	17			ABE	4.9-kb deletion
	18				human control
	19			KLF1	4.9-kb deletion
	20				human control
	21			CBE	4.9-kb deletion
	22				human control
	23			LRF 8C	4.9-kb deletion
	24				human control
	25	Unrelated samples			
	26				
	27	H_2O			4.9-kb deletion
	28				human control



Row	Number	Sample				
C	1	HD	Bone marrow	Mock		4.9-kb deletion
	2					human control
	3			CBE		4.9-kb deletion
	4					human control
	5			LRF 8C		4.9-kb deletion
	6					human control
	7			ABE		4.9-kb deletion
	8					human control
	9			KLF1		4.9-kb deletion
	10					human control
	11	SCD	Input	CFU-GM	Mock	4.9-kb deletion
	12					human control
	13				LRF 8C	4.9-kb deletion
	14					human control
	15				KLF1	4.9-kb deletion
	16					human control
	17		Bone marrow	Mock		4.9-kb deletion
	18					human control
	19			LRF 8C		4.9-kb deletion
	20					human control
	21			KLF1		4.9-kb deletion
	22					human control
	23			Mock		4.9-kb deletion
	24					human control
	25	Unrelated samples				
	26					
	27	H_2O				4.9-kb deletion
	28					human control

D	1	SCD	Bone marrow	Mock	4.9-kb deletion
	2				human control
	3			Mock	4.9-kb deletion
	4				human control
	5			Mock	4.9-kb deletion
	6				human control
	7			KLF1	4.9-kb deletion
	8				human control
	9			KLF1	4.9-kb deletion
	10				human control
	11			KLF1	4.9-kb deletion
	12				human control
	13			Mock	4.9-kb deletion
	14				human control
	15			KLF1	4.9-kb deletion
	16				human control
	17			KLF1	4.9-kb deletion
	18				human control
	19			LRF 8C	4.9-kb deletion
	20				human control
	21			Mock	4.9-kb deletion
	22				human control
	23	Unrelated samples			
	24				
	25	H_2O			4.9-kb deletion
	26				human control
	27	-			
	28	-			

Supplementary References

1. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
2. Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
3. Rosello, M. *et al.* Precise base editing for the in vivo study of developmental signaling and human pathologies in zebrafish. *eLife* **10**, e65552 (2021).
4. Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
5. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
6. Miller, S. M. *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**, 471–481 (2020).
7. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
8. Weber, L. *et al.* Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. *Sci. Adv.* **6**, eaay9392 (2020).
9. Knipping, F. *et al.* Disruption of HIV-1 co-receptors CCR5 and CXCR4 in primary human T cells and hematopoietic stem and progenitor cells using base editing. *Mol. Ther. J. Am. Soc. Gene Ther.* **30**, 130–144 (2022).
10. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168–e168 (2014).
11. Conant, D. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J.* **5**, 123–130 (2022).
12. Gaudelli, N. M. *et al.* Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).