UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE IN MEDICINA MOLECOLARE

Curriculum in Gene and Cell Therapy

UNCOVERING UPSIDES AND PITFALLS OF CAS, BASE AND PRIME EDITING IN HEMATOPOIETIC STEM/PROGENITOR CELLS

DoS: Prof. Luigi Naldini

Second Supervisor: Dr. Vikram Pattanayak

Tesi di DOTTORATO di RICERCA di Martina Fiumara

matr. 017135

Ciclo di dottorato XXXVI

SSD BIO11 / BIO18

Anno Accademico 2022/2023

CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il/la sottoscritto/*I*	Martina Fiumara
Matricola / *registration number*	017135
nat_ a/ *born at*	Fiesole (FI)
il/*on*	11/02/1995

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

UNCOVERING UPSIDES AND PITFALLS OF CAS, BASE AND PRIME EDITING IN HEMATOPOIETIC STEM/PROGENITOR CELLS

☒ AUTORIZZA la Consultazione della tesi / AUTHORIZES *the public release of the thesis*

☐ NON AUTORIZZA la Consultazione della tesi per …… mesi /DOES NOT AUTHORIZE *the public release of the thesis for …… months*

a partire dalla data di conseguimento del titolo e precisamente */ from the PhD thesis date, specifically*

Dal / *from* ……/……/…… Al / *to* ……/……/……

Poiché /*because*:

□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

□ ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

□ la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it and on its publication.*

Data /Date 23/10/2023 Firma /Signature Martina Fiumara

DECLARATION

This thesis has been:

- composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.

- has been written according to the editing guidelines approved by the University.

Permission to use images and other material covered by copyright has been sought and obtained (Italian legislative Decree no. 68/2003).

All the results presented here were obtained by myself, except for:

1) DNA sequencing at the target site, RNA-seq, clonal tracking and whole exome sequencing bioinformatic analyses which were performed by Dr. Stefano Beretta (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan)
2) Barcoded lentiviral vector construct which was designed and produced by Dr. Bernhard Gentner (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan), Dr. Matteo Maria Naldini (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan), Dr. Erika Zonari (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan) and Dr. Giacomo De Santis (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan).
3) Clonal tracking bioinformatic analyses which were performed by Dr. Stefano Beretta (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan) and Dr. Matteo Barcella (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan)
4) SMArT-3 experiments in the therapeutic configuration (Figure 34), which were performed by Dr. Daniele Canarutto (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan).

Other results in this PhD thesis were obtained in a joint effort:

1) Statistical analyses were performed in collaboration with Chiara Brombin (Centro Universitario di Statistica per le Scienze Biomediche) and Federica Cugnata (Centro Universitario di Statistica per le Scienze Biomediche (CUSSB))
2) SMArT experiments were performed together with Dr. Samuele Ferrari (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan) and Dr. Daniele Canarutto (SR-TIGET, IRCCS San Raffaele Scientific Institute, Milan).

Most of the studies on Base and Prime editing reported in this PhD thesis have now been published in a research article (on peer-reviewed scientific journal) where I am listed as a first author (Fiumara *et al*, 2023) and which I contributed to write together with Dr. Samuele Ferrari and Prof. Luigi Naldini. Thus, part of the text, mostly in the Results section, Figures of this thesis and Discussion section are shared with the cognate publication.

The edited cell selection studies have not been published yet and have been included in a patent application where I am listed as co-inventor (WO2020074729A1).

All sources of information are acknowledged by means of reference.

Acknowledgements

I would like to express my gratitude to Professor Luigi Naldini for his invaluable mentorship during my PhD and for the incredible opportunity he provided by allowing me to work in his laboratory. I had the privilege of being involved in various captivating projects under his guidance. I also extend my thanks for his meticulous review of my thesis. I am deeply appreciative of the opportunity to grow as both a scientist and an individual, and for the continuous inspiration I have received on this journey.

A special thanks goes to Samuele Ferrari who has been by my side throughout this entire voyage. Thank you for your unwavering support, for the scientific discussions, and for your assistance in designing and interpreting results. Collaborating and sharing my first publication as first author with you have been an honor. Your guidance and teachings have substantially contributed to my personal and professional growth. I am also grateful for your crucial input in revising my thesis.

I would like to thank all the past and present member of the Editing Team for the everyday work and support. Thanks Daniele Canarutto, Attya Omer-Javed, Luisa Albano, Chiara Gaddoni, Angelica Varesi, Gabriele Pedrazzani, Alessandra Weber, Valentina Vavassori and Aurelien Jacob, you are exceptional colleagues. It has been a pleasure working with all of you.

I would like to express my gratitude to Stefano Beretta for his invaluable contributions in performing bioinformatic analyses. Your hard work and dedication have been essential to the success of this project.

I am also grateful to all the members of Naldini's lab for their constant support. Without you, this journey would have been more challenging.

My thanks also go out to all our collaborators for their invaluable contributions, including Erika Zonari, Matteo Maria Naldini, Matteo Barcella, Bernhard Gentner, and Ivan Merelli.

Abstract

Hematopoietic stem and progenitor cells (HSPC) gene editing holds great promise for the treatment of inherited diseases by precise genetic engineering. Base and prime editing may provide safer and more precise genetic engineering as compared to nuclease-based approaches bypassing the dependence on DNA double strand breaks (DSBs). However, these systems are limited to correct single point mutations and/or introduce short sequence edits. Conversely, nuclease-based editing leveraging on homology directed repair (HDR) remains the treatment of choice when aiming to targeted integration of large, therapeutic DNA sequences. Despite promising results having been reported for base and prime editing in HSPC, little is known about adverse responses, on-target and genome-wide genotoxicity. Here, we performed a comparative side-by-side assessment of base editors (BE) and prime editor (PE) versus Cas9 nuclease for editing in HSPCs. While optimized timing and extent of editor expression allowed reaching the intended editing with high efficiency in long-term repopulating HSCs, all platforms induced multifaceted transcriptional responses that negatively impacted editing efficiency and/or clonogenic and repopulation capacity, albeit to a lesser extent than conventional Cas9. Induction of DNA DSBs and their genotoxic byproducts, such as long-range deletions and translocations, were less frequent compared to Cas9 but not abrogated by any of these platforms. Incomplete rather than full base excision repair inhibition aggravated these outcomes for the cytidine BE. Importantly, an unbiased genome-wide analysis uncovered a global impact of BEs on the mutational landscape of the hematopoietic graft. We observed an increased load and altered proportions of nucleotide variants, raising concerns for guide independent deamination due to the constitutive deaminase activity and the saturation or inhibition of endogenous repair pathways. In parallel, we developed strategies to enrich for HDR-edited HSPCs and reach high proportion of cells bearing the intended edit in the infused cell product. Selection by Means of Artificial Transactivator (SMArT) successfully enriched for HDR-edited HSPCs and purged out HSPCs bearing undesired, and potentially genotoxic, on-target DNA DSB byproducts. SMArT-edited HSPCs engrafted in immunodeficient mice and gave rise to fully edited hematopoietic grafts. Overall, the blueprint, set of metrics and innovative editing strategies described in this study will instruct careful and comprehensive evaluation of editing systems, and broaden applicability of HSPC gene editing toward clinical translation.

TABLE OF CONTENT

1 ACRONYMS AND ABBREVIATIONS

2A	Self-cleaving peptides
7-AAD	7-Amino-Actinomycin D
aa	Amino acids
AAV	Adeno-associated vectors
AAVS1	Adeno-associated virus site 1
ABE	Adenine base editor
ADA	Adenosine Deaminase
AdV	Adenoviral vector
ARCA	Anti-Reverse Cap Analog
ArT	Artificial trans activator
B2M	Beta-2 microglobulin
BAR	Barcode
BE	Base editor
BER	Base excision repair
BM	Bone marrow
bp	Base pairs
CB	Cord blood
CBE	Cytosine base editor
CFU-C	Colony-forming unit cell
CGD	Chronic Granulomatous Disease
CLPs	Common lymphoid progenitors
CMPs	Common myeloid progenitors
CMV	Cytomegalovirus
cPPT	Central polypurine tract sequence
CRISPR	Clustered regularly interspaced short palindromic repeats
DBD	DNA binding domain
dCas9	Dead Cas9
ddPCR	Droplet digital PCR
DDR	DNA damage response
dHJ	Double holliday junction
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
dNGFR	Tuncated low-affinity Nerve Growth Factor Receptor
DSB	Double strand breaks
eIF4G	Eukaryotic Initiation Factor 4G
EPCR	Endothelial protein receptor
eSpCas9	Enhanced specificity Staphylococcus pyogenes Cas9
FACS	Fluorescence activated cell sorting
FLT-3L	FMS-related tyrosine kinase 3 ligand

G-CSF	Granulocyte-colony stimulating factor
GMPs	Granulocyte-macrophage progenitor cells
GT	Gene therapy
guide RNA	gRNA
GvHD	Graft-versus-host disease
HBB	Human β-globin
HbF	Fetal hemoglobin
HBG	Human γ-globin
HDR	Homologous-directed repair
HJ	Holliday junction
HLAs	Human leukocyte antigens
HPLC	High performance liquid chromatography
HSC	Hematopoietic stem cells
HSCT	HSC transplantation
HSPCs	Hematopoietic stem and progenitor cell
IDLVs	Integrase defective lentiviral vectors
IFN	Interferon
IL	Interleukin
IL2RG	Interleukin-2-receptor γ chain
indels	Insertion or deletions
IRES	Internal ribosome entry sites
IT-HSC	Intermediate HSCs
kb	Kilobases
KO	Knock out
LIG4	DNA ligase IV
Lin	Lineage
LMPP	Lymphoid-primed multipotent progenitor
LNP	Lipid nanoparticles
LTC-IC	Long-term culture-initiating cells
LT-HSC	Long-term HSCs
LTR	Long terminal repeats
LV	Lentiviral vectors
MEP	Megakaryocyte-erythrocyte progenitor
minCMV	Minimal Cytomegalovirus promoter
MMEJ	Microhomology-mediated end joining
MMR	Mismatch Repair
mPB	Mobilized peripheral blood
MPPs	Multipotent progenitors
nCas9	Nickase Cas9
NER	Nucleotide excision repair
NHEJ	Non homologous end joining
NOD	Nonobese diabetic
opt	Optimized
PBS	Primer binding sequence

PE	Prime editor
pegRNA	Prime editor guide RNA
PGK	Phosphoglycerate kinase
Pol	Polymerase
polyA	Polyadenylation
PPP1R12C	Protein phosphatase 1 regulatory subunit 12C
PRRs	Pattern recognition receptors
RFI	Relative fluorescence intensity
RNP	Ribonucleoprotein
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase
RVD	Repeated Variable Diresidues
SCD	Sickle cell disease
SCF	Stem cell factor
SCID	Severe combined immunodeficiency
SCID-X1	X-linked severe combined immunodeficient
scRNA-seq	Single-cell RNA sequencing
SIN	Self-inactivating
SNV	Single nucleotide variant
SpCas9	Staphylococcus pyogenes Cas9
SpCas9-HF1	High fidelity Staphylococcus pyogenes Cas9
SPL	Spleen
SR1	StemRegenin 1
SSB	Single-strand breaks
ssODN	Single stranded oligodeoxynucleotide
ssRNA+	Single-strand positively oriented RNA
std	Standard
ST-HSC	Short-term HSCs
TadA*	TadA variant
TALE	Transcription activator-like effectors
TALEN	Transcriptional activator-like effector nucleases
TPO	Thrombopoietin
trugRNA	Truncated guide RNA
UG	Uracil glycosylase
UGI	Uracil glycosylase inhibitor
UTR	Untranslated region
VSV.G	Vesicular stomatitis virus glycoprotein
WAS	Wiskott-Aldrich syndrome
WES	Whole exome sequencing
WPRE	Woodchuck hepatitis virus post-transcriptional regulation element
WPRE	Woodchuck Hepatitis Virus Post-Transcriptional Regulatory Element

XLF	XRCC4-like factor
ZF	Zinc-finger
ZFN	Zinc-finger nucleases
β-thal	β-thalassemia
γRV	Gamma-retroviral vectors

2 LIST OF FIGURES AND TABLES

3 INTRODUCTION

3.1 Hematopoietic stem cell (HSC) biology

3.1.1 Hematopoietic hierarchy and hematopoiesis

Hematopoietic stem cells (HSC) are at the top of hematopoietic hierarchy and are defined by two main properties ensured through asymmetric division (Seita & Weissman, 2010):

(i) Self-renewal: the ability to sustain the stem cell reservoir during cell division.

(ii) Multipotency: the capacity to produce specialized cells that lose their stem-like properties and gradually undergo specialization, ultimately leading to the development of all mature blood cells that play crucial roles in processes, such as coagulation, oxygen transportation, tissue remodeling, and immune response.

The concept of universal progenitor for all blood cells was initially postulated at the early 20th century by the Russian biologist A. Maximow (Maximow, 1909) and became evident from studies on individuals exposed to lethal dose radiation, which survived upon transplantation of bone marrow (BM) or spleen (SPL) cells (Lorenz *et al*, 1951). A decade later, Till and McCulloch showed that stem cells harvested from the BM can give rise to multi-lineage progenitors (Till & McCulloch, 1961). Throughout childhood and adulthood, HSCs predominantly reside within hematopoietic niches in the BM, situated within the medullary cavity of long and flat bones. However, recent studies revealed the presence of circulating HSCs in different proportions throughout the life (Massberg *et al*, 2007). Hematopoiesis is the process responsible for blood cells formation and it takes place in the BM. Thanks to this process more than 10^{12} post-mitotic short-lived committed cells are daily produced in adults (Doulatov *et al*, 2012). A hierarchical organization guarantees a continuous blood cell regeneration, wherein a limited pool of stem cells give rise to all the differentiated blood cells (Boulais & Frenette, 2015). Several studies highlighted the heterogeneity of HSCs in humans, with cells bearing variable self-renewal and differentiation capacity and in different cell cycle states (Foudi *et al*, 2009; Wilson *et al*, 2008). The most primitive HSCs remain in a quiescent status (G0), which is supposed to be a way to prevent the possible errors that occur during division and could drive malignant transformation (Nakamura-Ishizu *et al*, 2014). Conversely, during differentiation HSCs enter in cell cycle and start to proliferate. Based on their self-

renewing and repopulation capacity, HSCs can be classified in three different subpopulations (Benveniste *et al*, 2010; Copley *et al*, 2012): long-term HSCs (LT-HSC), intermediate HSCs (IT-HSC) and short-term HSCs (ST-HSC) that progressively lose self-renewing potential and transit towards more differentiated cells. The immediate progeny of ST-HSCs are the multipotent progenitors (MPPs), which comprise the subpopulations MPP2, MPP3, and MPP4/lymphoid-primed multipotent progenitor (LMPP) (Zhang *et al*, 2018). MPPs are cells that have lost their self-renewal capacity but are able to proliferate and differentiate into different lineage-committed cells (Seita & Weissman, 2010). MMP2s and MMP3s follow a path of differentiation leading them to common myeloid progenitors (CMPs), whereas MMP4s/LMPPs embark on a differentiation journey towards common lymphoid progenitors (CLPs). CLPs represent oligopotent cells with the capacity to generate B cells and natural killer cells, while after migration to the thymus, they differentiate into T cells. CMPs instead give rise to megakaryocyte-erythrocyte progenitor (MEP) cells and granulocyte-macrophage progenitor cells (GMPs). MEPs contribute to the production of mature red blood cells and megakaryocytes, whereas GMPs give rise to monocytes, various types of granulocytes (including neutrophils, basophils, and eosinophils), and dendritic cells. The recent advancement of single-cell RNA sequencing (scRNA-seq) technologies, which provide an unbiased standpoint of the transcriptome, have provided novel insights into hematopoiesis and allowed to trace HSC origin, to reconstruct hematopoietic hierarchy and to understand HSC heterogeneity (Zhang *et al*, 2022). These studies highlighted a continuum of hematopoietic differentiation toward lineage specifications, as opposed to the stepwise commitment proposed in the classical hematopoietic tree (Laurenti & Göttgens, 2018). It is thus conceivable that LT-HSCs are heterogenous and that their fate is early defined during development. Indeed, single cell transplantation from HSC pool revealed that most of these cells are already primed with different lineage-biases (Carrelha *et al*, 2018). Lineage bias also correlate with age; indeed, aged HSC acquire a megakaryocytic bias and reduce their lymphoid and erythroid output (Grover *et al*, 2016).

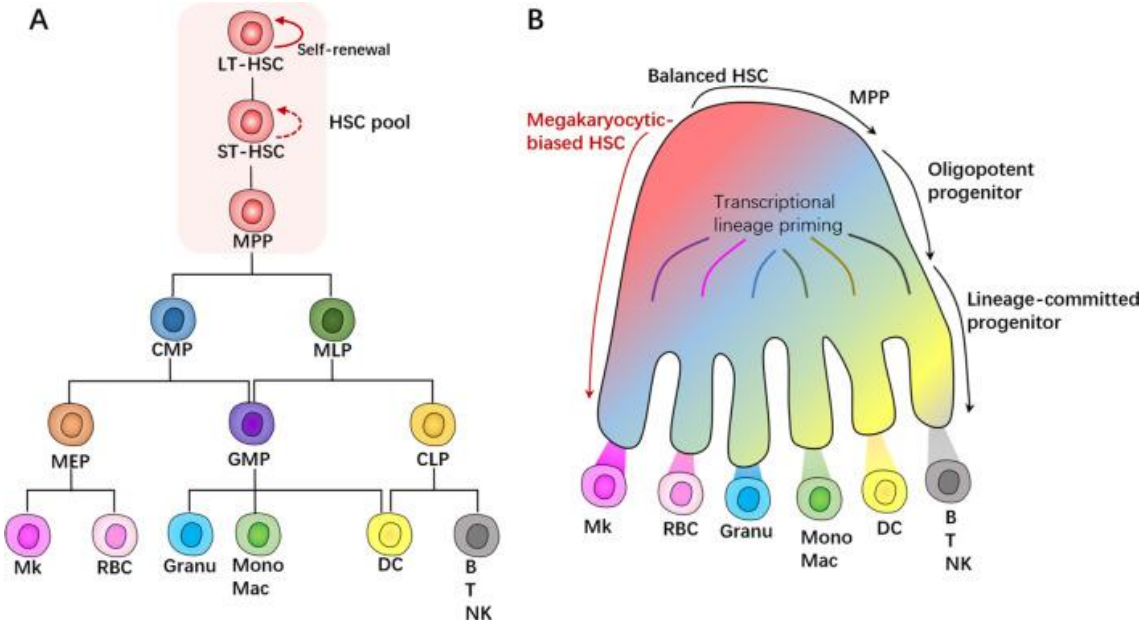


Figure 1. Schematic representation of hematopoietic hierarchy. *a, representation of the classical hematopoietic tree. b, representation of the hematopoietic hierarchy as result from single cell transcriptomic standpoint. Adapted from (Zhang* et al*, 2022).*

3.1.2 Phenotypic identification of HSPCs

Phenotypic identification of HSPCs (hematopoietic stem and progenitor cell) is achievable with antibodies tagged with fluorescent molecules that bind to specific antigens on the cell membrane, a technique known as fluorescence-activated cell sorting (FACS). The first surface marker able to identify a population enriched in hematopoietic multipotent and oligopotent progenitor cells was Cluster of Differentiation 34 (CD34), a glycoprotein that facilitates cell-to-cell adhesion through L-selectin (Civin *et al*, 1984). Few years later, it was demonstrated that transplanted CD34+ cells retain the ability to rebuild hematopoiesis after irradiation (Berenson *et al*, 1988). In the early 1990s, a Lin-CD34+CD90+ marker combination was introduced to identify a population able to establish multipotent long-term cultures and to generate hematopoietic progeny and ensure long-term multilineage reconstitution upon transplantation into immunodeficient mice (Baum *et al*, 1992; Weissman & Shizuru, 2008). While Lineage (Lin) markers pertain to molecules typically expressed by specific mature cell populations, CD90 (also known as Thy1) is a surface glycoprotein involved in cell-matrix interaction. For a more precise characterization of HSPCs, CD133 marker, a pentaspan transmembrane glycoprotein, whose expression is lost during progenitor differentiation was introduced

(Yin *et al*, 1997). To further enrich for HSC compartment, markers like CD38low/- and CD45RA- can also be employed, with CD38 (or cyclic ADP ribose hydrolase) and CD45RA (an isoform of protein tyrosine phosphatase receptor type C) progressively expressed during HSC differentiation (Bhatia *et al*, 1997; Hao *et al*, 1996). LT-HSCs are also characterized by the expression of CD49f which is lost in MPPs (Notta *et al*, 2011), and by the expression of the endothelial protein receptor (EPCR), which is not involved in HSC homing but is required for repopulation potential in primary and secondary recipients (Fares *et al*, 2017).

3.1.3 In vitro and in vivo systems to study human hematopoiesis

3.1.3.1 Clonogenic assays

The clonogenic assay is an *in vitro* method based on the ability of a single cell to growth up as a colony, and it allows to investigate human hematopoiesis. By plating HSPCs in semi-solid media, such as soft agar, methylcellulose, or plasma gel fibrin clots, it is possible to restrict cell movement and facilitate differentiation into cell colonies *in vitro*. In the presence of colony-stimulating factors, only progenitor cells plated in semi-solid materials are capable of growth and can give rise to mature hematopoietic colonies known as CFU-C (colony-forming unit cells) in approximately two weeks. Consequently, the number of generated colonies provides an estimation of the number of individual progenitors initially plated. Moreover, morphological criteria allow to count and distinguish between myeloid and erythroid colonies. Since even the more committed human HSPCs partially retain some colony forming potential, long-term culture-initiating cells (LTC-IC) assays were developed for *in vitro* functional investigation of more primitive HSCs. However, the relationship between long-term culture-initiating cells and HSCs remains unclear, suggesting the necessity of *in vivo* models to comprehensively study the repopulation potential of LT-HSPCs.

3.1.3.2 Xenograft models

Several *in vivo* mouse models have been developed to study and recapitulate the hematopoietic reconstitution by LT-HSPCs following BM transplantation. In a first attempt hematopoietic progenitors derived from BM were transplanted into CB17-severe combined immunodeficient (SCID) mice (Kamel-Reid & Dick, 1988). However, cross-reactivity between murine and human myeloid growth factors prevented myeloid

engraftment. To stimulate myeloid engraftment, IL-3, GM-CSF, and SCF cytokines were infused into CB17-SCID mice (Lapidot *et al*, 1992). Cells able to generate both lymphoid and myeloid progeny in the xenotransplant setting were termed SCID-repopulating cells. However, this xenograft model achieved only limited and transient engraftment due to the generation of mouse B and T cells and the presence of high levels of host natural killer cells. In 1995, higher levels of human engraftment were achieved by introducing the SCID mutation into the nonobese diabetic (NOD) genetic background which lacks natural killer cell function and exhibit deficiencies in macrophages and the complement system (Greiner *et al*, 1995). The resulting NOD/SCID model led to increased chimerism upon HSC transplantation, but challenges such as low human B and T cell maturation and thymic lymphoma occurrence prevented long-term studies. A pivotal advancement occurred when immunodeficient mice homozygous for targeted mutations in the interleukin-2-receptor γ chain (*IL2RG*) were crossed with NOD/Shi-SCID, resulting in NOG mice, and with NOD/LtSz-SCID, resulting in NSG mice. NOG and NSG mice lack B, T, and NK cells and provide robust support for high levels of long-term engraftment (Goyama *et al*, 2015). However, these mouse strains display progressive reduction of human engraftment from six months post-transplantation due to the failure of interaction between mouse cytokines and human receptors. To overcome this hurdle, various strategies have been devised:

(i) Supplement with human recombinant proteins;

(ii) *In vivo* lentiviral gene delivery of human cytokines;

(iii) Hydrodynamic injection of plasmid DNA encoding for human cytokines (Rongvaux *et al,* 2013);

(iv) Transgenic NSGS mice for human cytokines SCF, GM-CSF, and IL-3 (Wunderlich *et al,* 2018);

(v) Knock-in MISTRG mice in which genes encoding for human cytokines M-CSF, human IL-3 and GM-CSF replace the mouse ones. In addition, MISTRG mice carry a transgene encoding the human SIRPα in a bacterial artificial chromosome. SIRP1α binds CD47 that is constitutively expressed by human cells, avoiding phagocytosis of engrafted human cells by the murine counterpart (Rongvaux *et al*, 2014).

However, the need of sublethal irradiation prior to transplantation and the inability to sustain long-term multilineage reconstitution in all proposed mouse models have recently encouraged the development of the mouse strains NSGW41 and NBSGW, both of which circumvent the preconditioning requirement thanks to a loss of function mutation in cKit which impairs murine hematopoiesis (Adigbli *et al*, 2021; McIntosh *et al*, 2015). Furthermore, to evaluate the interaction among human HSC and the BM niche advanced *in vivo* model have been developed based on the use of ossicles (Abarrategi *et al*, 2017). Ossicles consist of scaffolds coated with human mesenchymal stromal cells that can be implanted in mice to create a humanized bone tissue, which support HSPC engraftment.

3.1.4 HSPC based therapies for inherited diseases

The evolving understanding of hematopoiesis has paved the way for significant advancements in HSPC transplantation (HSCT) within clinical practice, addressing a spectrum of both malignant and non-malignant conditions. The primary objective of HSCT is to replace patient's defective cells with healthy ones sourced from a donor. Depending on the origin of HSPCs, transplantation can be defined as:

- Autologous: transplanted cells are derived from the same individual;
- Syngeneic: the donor is an identical twin;
- Allogeneic: cells are collected from an unrelated individual.

Alternative sources for HSPC harvesting are:

(i) Isolation from cord blood (CB);

(ii) Harvesting of BM from iliac crests;

(iii) Leukapheresis of HSPCs from peripheral blood after mobilization (mPB).

Since the latter approach is a less invasive collection method in comparison to BM aspiration from the iliac crests, it represents the most widely used in clinical practice (Canarutto *et al*, 2021). While initial mobilization protocols involved chemotherapy alone (Stiff *et al*, 1983), nowadays the most commonly used agent is granulocyte-colony stimulating factor (G-CSF), which has garnered approval from regulatory agencies (Publicover *et al*, 2013). Recombinant human G-CSF, produced in *Escherichia coli*, triggers the expansion of granulocytes and the modulation of adhesion molecules, both in protease-dependent and independent manners. This leads to the disruption

CXCL12/CXCR4 binding and HSPCs egression from BM (Altuntaş & Korkmaz, 2017). To increase the number of collected cells, G-CSF can be combined with Plerixafor, a reversible inhibitor of CXCL12/CXCR4 binding (Matthys *et al*, 2001). The combination of the two drugs was approved for clinical application (DiPersio *et al*, 2009b, 2009a). Furthermore, Bio5192 has been used in combination with G-CSF and Plerixafor to destroy VCAM-1/VLA-4 axis, which is responsible for HSPCs homing in the BM, resulting in 17-fold increase in mobilization compared to G-CSF only treatment in a preclinical study (Ramirez *et al*, 2009). To achieve faster mobilization compared to the standard G-CSF regimen, plerixafor was recently combined with GROβ, a CXCR2 agonist in a preclinical study (Hoggatt *et al*, 2018).

Independently from HSPC origin, recipients undergo partial or full myeloablative conditioning regimen in order to make space in the BM niche for the engraftment of donor cells prior to HSCT (Gyurkocza & Sandmaier, 2014). Conditioning regimens typically consist of chemotherapy or radiotherapy treatments which expose the recipient to short-term and long-term side effects. On one hand, in the short-term those treatments induce immune suppression thus increasing the risk of infections; on the other hand, conditioning regimens alter the BM niche architecture impairing the donor cells engraftment. Only for some specific pathological conditions in which endogenous HSPCs are fewer in the BM niche, e,g, Fanconi anemia, conditioning regimen can be avoided (Río *et al*, 2019a).

3.1.4.1 Allogeneic HSCT

In allogeneic transplantation settings, the genetic variability between the donor and recipient plays a crucial role. Inadequate conditioning of the recipient may lead to graft rejection due to innate and adaptive responses against the donor cells (Olsson *et al*, 2013; Hutt, 2018; Lowsky & Messner, 2016). To prevent rejection, immunosuppressive drugs are typically administered prior to HSC transplantation (Hamilton, 2018). Conversely, donor's lymphocytes may trigger an immunological reaction against the host tissues, known as graft-versus-host disease (GvHD). Biologically, GvHD is triggered by the histocompatibility complex, characterized by a distinct set of human leukocyte antigens (HLAs), which play a role in the immune system's identification of self and non-self (Lowsky & Messner, 2016).

3.1.4.2 Autologous HSCT

Autologous HSCT addresses the challenges related to rejection and GvHD and finds its application for the treatment of hematological disorders, autoimmune diseases, and solid tumors resistant to standard therapies (Snowden, 2016). Autologous HSCT shows reduced risk of infection and prompter hematopoietic and immunological recovery post-transplantation compared to allogeneic HSCT (Tucci *et al*, 2021).

3.2 HSPC gene therapy

Gene therapy (GT) enables the insertion of a DNA fragment into a defective cell, either to substitute the impaired function or to provide the cells with a new function able to revert the pathological progression (Verma & Weitzman, 2005). In HSPC GT, patients' cells are harvested, *ex vivo* modified and reinfused back after conditioning regimen, which is meant to pre-empty the BM niche and make space for engineered cells to engraft. Although either non-viral or viral platforms can be employed to transfer genetic material, in GT approaches, a therapeutic cassette is often delivered for therapeutic purposes through viral vector which can be either non-integrating or integrating depending on the viruses from which they have been evolved. Non-integrating vectors, such as adenoviral (AdV) or adeno-associated vectors (AAVs) (McCarty *et al*, 2004), are widely used for gene delivery since they exhibit tropism for specific cell types and high transduction efficiency. Whereas non-integrating vectors persist in post-mitotic tissues, their repeated administration is required to achieve stable long-term expression of the therapeutic gene in proliferating cells because of particles dilution upon cell division. To guarantee life-long expression of the therapeutic cassette while avoiding multiple administration, which may lead to cell mediated immune response against the vector (Mueller & Flotte, 2008), integrating vectors represent the standard of choice. Among integrating vectors, both gamma-retroviral vectors (γRV) (Maier *et al*, 2010) and lentiviral vectors (LV) (Naldini *et al*, 2016) are used in GT context, with the latter being the most successful platform. One of the most relevant advantage of LVs compared to γRVs is that they can transduce not only proliferating cells but also non-dividing cells taking advantage of the nuclear pore transport system as shuttle for nuclear entry (Naldini, 1998). Furthermore, LVs are not affected by epigenetic silencing targeting the vector sequence and thus affecting most integrated vectors independently from the insertion sites, while such events have been reported for γRVs in some cell types (Challita & Kohn, 2006).

3.2.1 Lentiviral vector

LVs are derived from Lentivirus which consist of a diploid viral genome composed by two copies of a single-strand positively oriented RNA ($ssRNA^+$) with length ranging from 7 to 10 kilobases (kb) surrounded by a truncated cone-shaped capsid and by an envelope. Lentiviruses infect cells through binding and fusion between envelope glycoproteins and receptor in the host cell membrane. Upon capsid uncoating, $ssRNA^+$ is reverse transcribed to dsDNA and integrated in the host cell genome in a semi-random manner (Lewinski & Bushman, 2005). The integrated viral genome is then transcribed to produce viral proteins and the $ssRNA^+$ genome, ultimately leading to the assembly and release of mature virions that can infect other cells. Viral genome consists of structural genes (i) *gag*, *pol* and *env*, (ii) genes involved in replication *Tat*, *Rev* (iii) four accessory genes *vif*, *vpr*, *vpu* e *nef* and (iv) Long terminal repeats (LTRs) at 5' and 3' terminal necessary for reverse-transcription, integration, and regulation of expression. In the context of GT, LVs have been evolved to minimize the risk of recombination and formation of a replication competent virus (Vigna & Naldini, 2000). Third generation LVs are obtained by transfection of four different plasmids in producer cells:

- The envelope construct contains a strong promoter (cytomegalovirus, CMV) driving expression of the vesicular stomatitis virus glycoprotein (VSV.G), which confers broad LV tropism;
- The packaging construct contains the same CMV promoter to express *gag* and *pol*;
- The rev construct drives synthesis of the rev protein, which acts in trans on the packaging construct;
- The self-inactivating (SIN) transfer construct. This construct contains modified LTRs that lack the U3 region, the packaging signal, an internal promoter and the transgene. U3 regions contain viral enhancer and promoter sequences, therefore their deletion causes transcriptional inactivation of LTRs. The deletion in U3 region is crucial to improve biosafety of LVs, preventing the risk of oncogene activation due to a possible enhancer effect on neighboring genes. Importantly, the U3 activity is replaced by a strong constitutive promoter derived from Rous Sarcoma Virus (RSV) or human Cytomegalovirus (CMV), thus abolishing Tat-transcriptional dependence in vector-producer cells. Therefore, in these advanced LVs, transgene expression is driven by an internal promoter. Furthermore,

improved SIN transfer vectors also contain the woodchuck hepatitis virus post-transcriptional regulation element (WPRE), which promotes polyadenylation improving at post-transcriptional level transgene expression, and the central polypurine tract sequence (cPPT), which enhances nuclear translocation of the vector genome.

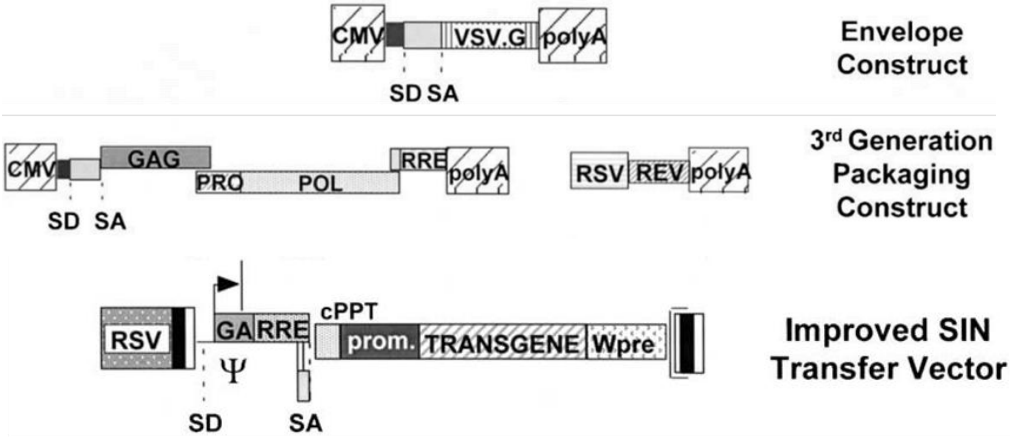


Figure 2. Schematic representation of the four plasmids transfected in producer cells for the generation of LVs. *Adapted from (Vigna & Naldini, 2000)*

3.2.2 Challenges of integrating vectors

The two major challenges attributed to the use of integrating vectors are insertional mutagenesis and regulation of transgene expression.

3.2.2.1 Insertional mutagenesis

Whenever a vector integrates its genetic material into a host cell's genome, the outcome can vary from being neutral to potentially causing genotoxicity, depending on the location in which the event occurs. If the integration happens in a non-transcribed and non-regulatory region the event can be neutral, vice versa if the genetic information is affected the event might be mutagenic. One concern arises from the strong enhancer and promoter sequence embedded into active LTRs, which can activate neighboring genes and potentially induce cell transformation (Naldini, 2011). Indeed, last generation of LVs contains modified LTR that lack enhancer and promoter region. Both RVs and LVs integrate with semi-random pattern (Bushman *et al*, 2005), with the formers showing propensity for regions involved in the transcriptional regulation by RNA polymerase II, such as promoters, transcriptional binding sites, DNase-I hypersensitive sites and

methylated CpG islands, and the latter showing propensity for bodies of genes and regions downstream of the transcription start site (Roth *et al*, 2011; Cavazza *et al*, 2013).

3.2.2.2 Unregulated expression of the transgene

One key aspect in GT is the extent of expression of the therapeutic transgene. When ubiquitous transgene expression is needed, promoters of "housekeeping" genes such as phosphoglycerate kinase (PGK) and Elongation Factor 1 α (EF1a) are often introduced in the transfer vector. Conversely, when aiming to restrict transgene expression to specific lineage, lineage specific promoters instead of ubiquitous ones are preferred (Chad *et al*, 2000). MicroRNA target sequence can be added in the 3' UTR region of the transgene to specifically inactivate its expression in cells expressing the microRNA (Brown *et al*, 2007; Gentner *et al*, 2009). Although these strategies enable some degree of gene expression control, precisely reproducing the endogenous regulation of a replaced gene remains challenging. This limitation discourages the application of viral vector mediated gene addition strategies in diseases requiring faithful, physiological regulation of the therapeutic transgene for safe and effective correction.

3.2.3 HSPC GT applications

Initial clinical HSPC GT studies leveraged on γRV to treat Adenosine Deaminase (ADA) Severe Combined Immunodeficiency (SCID) by permanently integrating into the HSPC genome the corrective form of the defective enzyme. Years of robust safety and efficacy clinical data led to market approval of the first ex vivo GT product with the commercial name of Strimvelis™ for the treatment of ADA-SCID (Cicalese *et al*, 2016). Despite the remarkable results in terms of efficiency, relevant safety concerns surfaced in early γRV trials for the treatment of X-linked SCID (SCID-X1), a genetic inborn error of immunity caused by mutation in the *IL2RG* gene. Some patients developed leukemia due to vector insertion near a proto-oncogene, leading to overexpression of genes promoting the development of acute lymphocytic leukemia (Hacein-Bey-Abina *et al*, 2003, 2008). Similar adverse events caused by γRV insertional mutagenesis were reported in patients affected by X-linked Chronic Granulomatous Disease (CGD) and Wiskott-Aldrich syndrome (WAS) (Ott *et al*, 2006; Braun *et al*, 2014). Thanks to their improved design and intrinsic biological properties, LVs became the preferred choice for HSPC GT (Vigna & Naldini, 2000). Nowadays, LVs find application in different hematological diseases

such as primary immunodeficiencies (De Ravin *et al*, 2016; Kohn *et al*, 2020, 2021; Ferrua *et al*, 2019) bone marrow failure (Río *et al*, 2019b), metabolic diseases (Sessa *et al*, 2016; Gentner *et al*, 2021) and hemoglobinopathies (Marktel *et al*, 2019; Esrick *et al*, 2021; Locatelli *et al*, 2022), achieving outstanding results in terms of safety and efficacy so far. Nevertheless, the potential drawbacks of GT, such as insertional mutagenesis events and non-physiological transgene expression, still require careful evaluation tailored to each specific application.

3.3 DNA repair machinery

Eukaryotic cells are daily exposed to DNA damage which introduces mutations that needs to be promptly repaired in somatic cells to ensure survival and avoid accumulation of mutations which may lead to loss of normal function and regulation, and to transformation (Chatterjee & Walker, 2017). DNA damage may originate from structural and chemical modifications induced by endogenous or exogenous agents, and from DNA polymerase involved in cell replication that intrinsically produce errors, albeit to very low extent. To guarantee the repair of such events, cells deploy complex machineries, termed DNA repair pathways, assigned to restore the original sequence upon specific damages. When damages cannot be fixed, cells activate transcriptional program leading to cell death or senescence. DNA damage may occur at one or both genomic strands. In the former case, the complementary strand can be used as template for repair; in the latter, either some genetic material is lost, or the sister chromatid is used to faithfully restore the native sequence.

3.3.1 Single strand DNA repair pathways

In eukaryotic cells, single strand DNA repair pathways play a crucial role in maintaining normal DNA synthesis, genomic stability, and integrity (Wang *et al*, 2023a). These processes are essential for various cellular functions, including cell proliferation, differentiation, cell cycle regulation, apoptosis, as well as for the development of tissues and organs. The major pathways involved in DNA SSB repair are base excision repair (BER), nucleotide excision repair (NER) and Mismatch Repair (MMR). BER is responsible for repairing deamination, methylation, alkylation, and oxidative damage to a single base, NER fixes damages involving multiple bases, which cause helix distortion,

while MMR detects and corrects structural alteration and errors occurring during DNA replication, repair, and recombination.

3.3.1.1 Base excision repair (BER) pathway

BER is specialized in recognizing and fixing single base damage that causes mispairing between the two genomic strand and may lead to mutagenesis during replication. Independently from the type of damaged base, BER works in three steps: (i) lesion recognition, (ii) excision of the mutated nucleotide, (iii) resynthesis (Lee & Kang, 2019a). The first step involves DNA glycosylases which have different substrate specificity and recognize the damaged nucleotide. After lesion recognition, DNA glycosylases induce the displacement of the injured base from the DNA helix and cleave an N-glycosidic bond, resulting in the liberation of a free base and the formation of an abasic site. Subsequently, this site undergoes further processing by AP endonuclease 1 (APE1), which cleaves the DNA backbone on the 5' side of the abasic site. This cleavage results in the formation of nick with a 3'-hydroxyl group and a 5'-2-deoxyribose-5'-phosphate (5'-dRP) residue. DNA polymerase β (Polβ) exploits the 3'-hydroxyl group to bridge the gap through template-directed synthesis using the complementary strand as template.

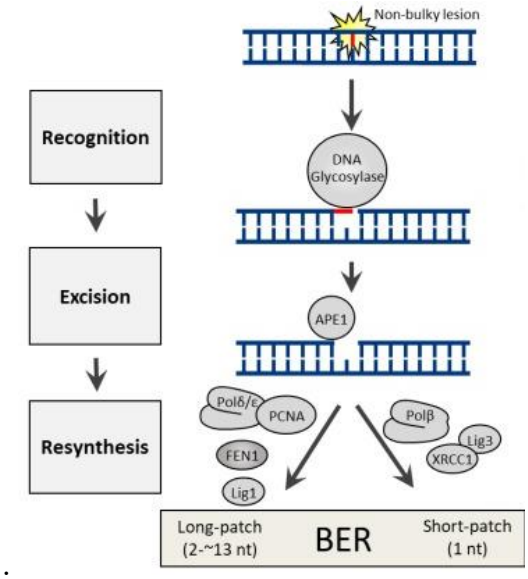


Figure 3. Schematic representation of the BER pathway. *Adapted from (Lee & Kang, 2019b)*

3.3.1.2 Nucleotide excision repair (NER) pathway

While BER starts with the specific recognition of the mutated base DNA glycosylase mediated, NER does not recognize the specific mutation but rather is activated by the presence of helix distortions (Wang *et al*, 2023b). NER is primarily involved in the resolution of pyrimidine dimers and other lesions induced by radiations. Once the damage has been recognized, nicks are introduced upstream and downstream the mutated region by endonucleases which create a ~20-30 bases gap on the mutated strand. The gap is then restored by DNA polymerase using the complementary strand as template.

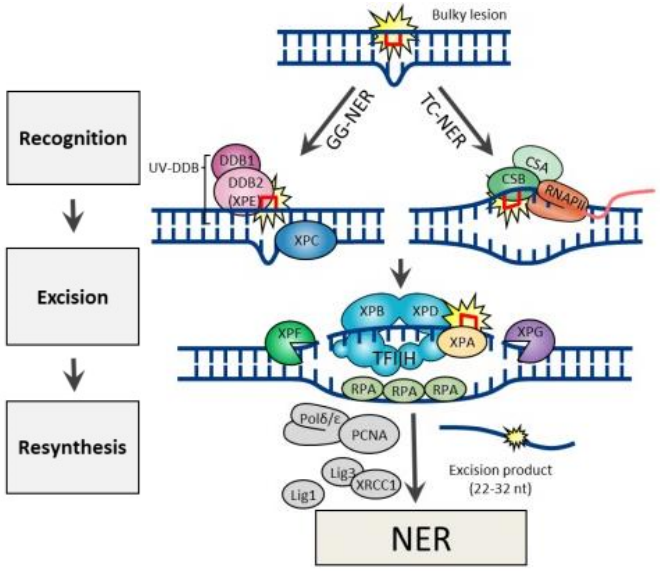


Figure 4. Schematic representation of the NER pathway. *Adapted from (Lee & Kang, 2019b)*

3.3.1.3 Mismatch repair (MMR) pathway

In eukaryotic cells, the MMR pathway begins with the recognition of mismatches by the MutSα and MutSβ heterodimeric complexes. Once a mismatch is detected, these complexes recruit MutLα, subsequently activating the endonuclease activity of MutH, which nicks the DNA strand. This nick facilitates the excision of the erroneous DNA strand containing the mismatch, a process mediated by EXO1. Following this excision, DNA polymerases initiates the resynthesis of the excised DNA segment (Kunkel & Erie, 2015). During DNA replication high fidelity polymerases δ and ε are mainly recruited to repristinate the original sequence; conversely, during specific physiological events, such

as immunoglobulin class-switching and somatic hypermutation, error-prone polymerases are recruited and generate genetic diversity (Li, 2007).

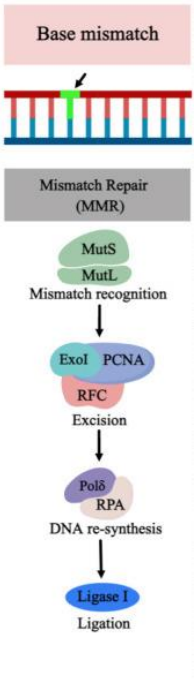


Figure 5. Schematic representation of the MMR pathway. *Adapted from (Wang* et al*, 2023a)*

3.3.2 Double strand repair pathways

DNA double strand breaks (DSBs) represents the most dangerous genomic event since they may cause gene disruption and chromosomal loss or translocations. The choice of the pathway for DNA DSB repair mostly depends on the phase of the cell cycle during which the injury occurred. In the G0/G1 phases, when a cell experiences damage, the broken ends are stitched by inserting or deleting some bases through non homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ). Conversely, if cells are in the S/G2 phases, the high fidelity homologous-directed repair (HDR) mechanism comes into play, employing a DNA template like the sister chromatid to accurately restore the correct sequence. The choice of repair is primarily regulated by two tumor suppressors: 53BP1 and BRCA1. During the G1 phase, 53BP1 promotes NHEJ-mediated processing of DNA DSB ends (Bunting *et al*, 2010), conversely, during S/G2 phases, BRCA1 promotes HDR (Escribano-Díaz *et al*, 2013).

3.3.2.1 Non-Homologous End Joining (NHEJ) pathway

NHEJ is the primary cellular mechanism for DNA DSB repair (Burma *et al*, 2006). NHEJ is an error-prone repair pathway that install small nucleotide insertions or deletions at the DSB site and is highly relevant to fix damages caused by ionizing radiation, DNA replication errors, reactive oxygen species, and in the context of the physiological V(D)J recombination process (Chang *et al*, 2017). The NHEJ pathway starts with the recognition of the DSB by the Ku70-Ku80 heterodimer, with each Ku dimer binding to one end of the DNA DSB (Britton, Coates, and Jackson 2013). At this juncture, the Ku70-Ku80 dimers recruits other NHEJ proteins, including the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), DNA ligase IV (LIG4), and its associated components XRCC4, XRCC4-like factor (XLF), and XLF (a paralogue of XRCC4, known as PAXX) (Gottlieb & Jackson, 1993; McElhinny *et al*, 2000; Ochi *et al*, 2015). A long-range synapsis configuration is formed between Ku70-Ku80 and DNA-PKcs, followed by a short-range synapsis, where the two ends are joined trough the actions of XLF, non-enzymatic XRCC4-LIG4 activities, and DNA-PKcs (Blackford & Jackson, 2017; Scully *et al*, 2019b). Concurrently, the recruitment of the nuclease Artemis and DNA polymerases Pol λ and Pol μ ensures the processing and removal of damaged ends. This dynamic process, characterized by alternating between long-range and short-range synapsis, may ultimately lead to XXCR4-LIG4-mediated ligation and disengagement of the NHEJ machinery. NHEJ can be engaged throughout the cell cycle.

3.3.2.2 Microhomology-Mediated end joining (MMEJ)

MMEJ has been described as an alternative mechanism to NHEJ, in which instead of the activation of classical NHEJ components, 5-25 nucleotides of microhomology sequences are used to repair the break (Chang *et al*, 2017). The MMEJ pathway is composed by four steps: (i) DNA DSB free ends are resected; (ii) the two exposed filaments are annealed by microhomology regions which results in two 3' overhangs flaps; (iii) 3' flaps are removed; (iv) DNA polymerase and ligase resolve the break (Wang & Xu, 2017). MMEJ is available only during S-G2 phases of the cell cycle.

3.3.2.3 Homology directed repair (HDR) pathway

HDR-mediated repair represents the second major pathway for resolving DNA DSB (Scully *et al*, 2019a). Beyond sequence similarity, physical cohesion and genomic

alignment between the target sequence and the repair template are also recognized as factors that facilitate HDR (Kadyk & Hartwell, 1992; Scully *et al*, 2019a). HDR may be preferred to NHEJ for DNA DSB repair in S and G2 phases, particularly if high copies of homologous sequences are available. Mechanistically, the DNA DSB sensor MRN, comprising three subunits (MRE11, RAD50, and NBS1), creates a nick at the 5' strand, extending up to 300 nucleotides upstream of the break site, and activates ATM (Garcia et al., 2011). CtBP-interacting protein (CtIP) is necessary to facilitate "short-range" nucleotide resection, mediated by the 3'-5' endonuclease activity of MRE11. Furthermore, the recruitment of EXO1, BLM, and DNA2 is essential for driving "long-range" resection via 5'-3' activity, displacing the Ku70-Ku80 heterodimers bound to the DNA ends. Subsequently, the RPA complex, composed of RPA1-RPA2 and RPA3, binds to the newly formed single-stranded DNA (ssDNA) to prevent secondary structures and unwanted interactions with other ssDNA sequences. For homologous recombination, other HDR mediators (BRCA2, PALB2, BRCA1-BARD1) must displace RPA, allowing RAD51 to bind to the ssDNA and form a nucleoprotein filament (San Filippo *et al*, 2008). This filament facilitates the search for homology sequences through duplex DNA invasion, promoting base-pairing between the invading strand and its complementary counterpart in the invaded DNA (Chen *et al*, 2008). BRCA1-BARD1 plays an important role in this process (Zhao et al. 2017). When the invading strand and its complementary counterpart are matched, the synapse is stabilized and the non-paired strand is relocated to form a displacement-loop (D-loop) through ATP hydrolysis and RAD51 dissociation (Van der heijden *et al*, 2007). Next, the free 3' terminus of the invading strand acts as a primer for DNA Pol δ, which extends the filament using the invaded DNA as a homologous template for DNA repair (Maloisel *et al*, 2008). Various pathways have been reported for synapse resolution. For instance, somatic cells typically prefer the non-crossover synthesis-dependent strand annealing mechanism. In non-crossover synthesis-dependent strand annealing mechanism, only one end of the DSB is guided by RAD51 for duplex DNA invasion, while the other end is passively resected. Notably, this mechanism is considered a non-crossover pathway, as it does not generate Holliday junctions (HJ) (Pâques & Haber, 1999). In contrast, meiotic cells use both ends of the DSB for strand invasion, potentially forming a double HJ (dHJ). In this case, either non-crossover or crossover mechanisms can be employed for dHJ resolution, with the latter contributing to genetic

diversity (Bizard & Hickson, 2014; Pâques & Haber, 1999). In cases of displacement failures, the error prone mechanisms long-tract gene conversion or break-induced replication may occur (Malkova & Ira, 2013).

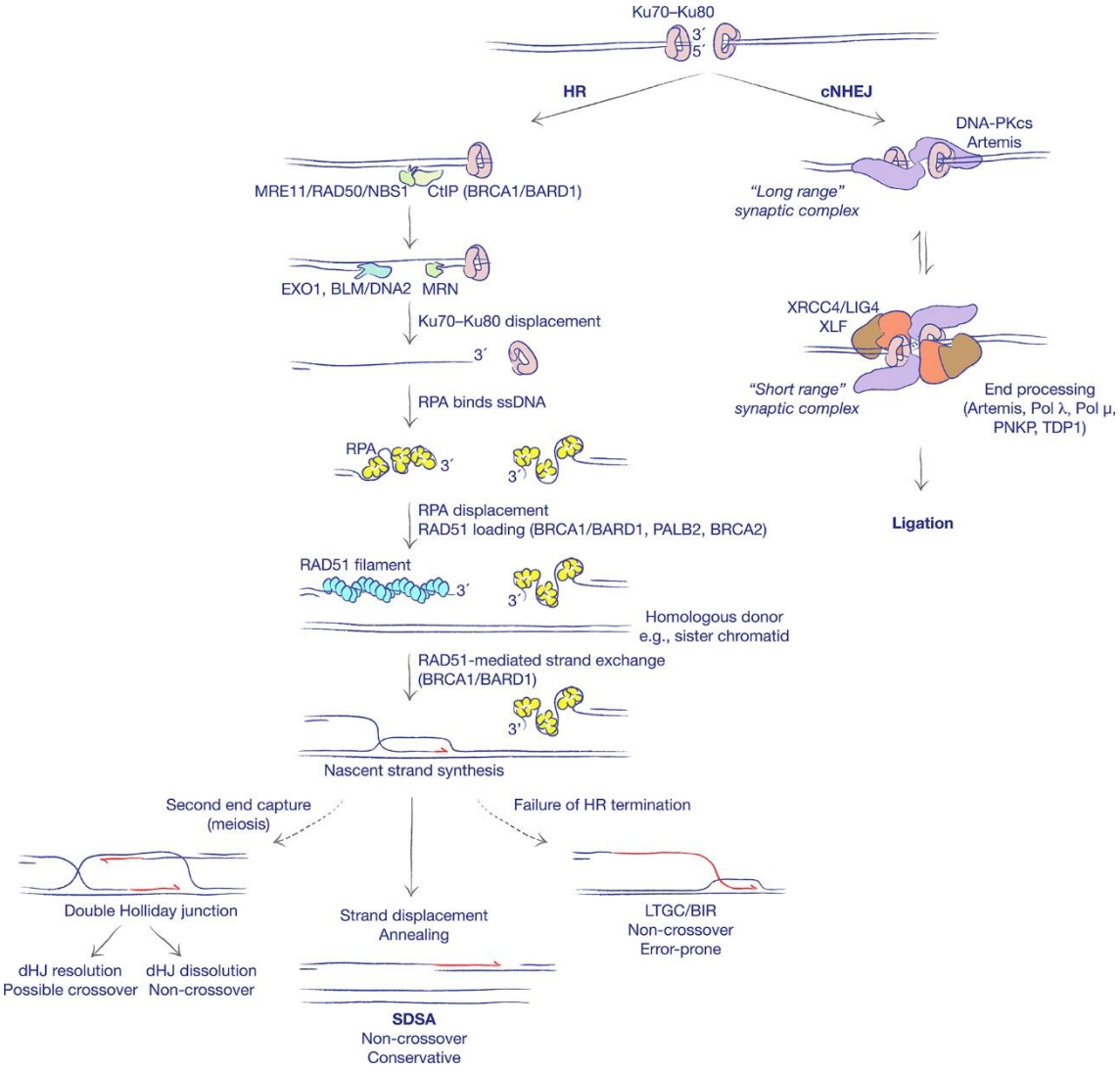


Figure 6. Schematic representation of NHEJ and HDR repair pathways. *Adapted from (Scully* et al*, 2019b). Permission to reproduce from Nature Reviews Molecular Cell Biology under license number 564146090128.*

3.4 Gene editing

Gene editing is a molecular biology technique that allows to precisely modify DNA, typically by adding, deleting, or substituting specific DNA sequences at targeted locations within the genome. Gene editing holds the promise for the treatment of inherited diseases while mitigating the risk related to genome-wide insertional mutagenesis and uncontrolled expression of the transgene, issues that emerged with viral vector mediated

gene addition therapies, and particularly with γRV. Gene editing components are (i) a nuclease directed to a specific target region, and eventually (ii) a donor template carrying homology arms (HA) for the DNA DSB to favor HDR.

3.4.1 Programmable nuclease

Nuclease-based gene editing exploits programmable chimeric molecules delivering a DNA DSB at the target genomic locus of interest. Over the past two decades, different chimeric molecules have been exploited to introduce DSB in the genome, including zinc-finger nucleases (ZFN) (Urnov *et al*, 2005; Lombardo *et al*, 2007) and transcriptional activator-like effector nucleases (TALEN) (Becker & Boch, 2021). However, the discovery of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas systems as adaptable tools to engineer the eukaryotic genome was an absolute breakthrough for the gene editing field.

3.4.1.1 Zinc finger nucleases (ZFNs)

ZFNs are chimeric protein composed by multiple ZF DNA binding domain (DBD) and a catalytic domain derived from the C-terminus of the restriction enzyme FokI, found in *Flavobacterium okeanokoites* (Handel & Cathomen, 2011). Each ZF DBD is composed by ~30 amino acids and recognizes 3-4 bases in the genome. From 3 to 6 DBD are linked through flexible linkers thus recognizing 9-18 nucleotides in the genome. Two ZFN monomers, which recognize the two opposite strands upstream and downstream the target site, are required to introduce a DNA DSB since the FolkI catalytic domain works as a dimer. Despite their optimization is challenging and time consuming, ZFNs remain an efficient tool for targeted gene editing and have been also used in the context of HSPCs (Genovese *et al*, 2014).

3.4.1.2 Transcription activator-like effector nucleases (TALENs)

Transcription activator-like effectors (TALE) are proteins secreted by a bacterium of genus *Xanthomonas*. Secreted TALEs can bind specific sequences in host cell, thus altering host gene transcription to promote bacterial infection (Boch & Bonas, 2010). An array of TALE DBD can be fused with FokI nuclease-domain to create the TALE nuclease monomer. Each TALE in the array is able to recognize a specific nucleotide in the target sequence through key hypervariable residues in position 12 and 13, called Repeated Variable Diresidues (RVDs) (Moscou & Bogdanove, 2009). Indeed, a simple code

associates 4 RVDs to the 4 nucleotides (Bogdanove & Voytas, 2011). As previously described for ZFN, two TALEN monomers are required to allow FokI nuclease dimerization which leads to DNA DSB introduction. Even though TALEN assembly results easier compared to ZFNs, the length of each repeat, the repetitiveness of DBD and the high homology among modules represent the major challenges for their use.

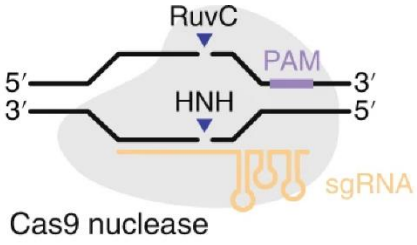
3.4.1.3 CRISPR/Cas9 nuclease

In nature, CRISPR/Cas systems operate within bacteria and archaea as a form of adaptive immunity against plasmids and viruses (Barrangou *et al*, 2007; Brouns *et al*, 2008). During infection, bacteria and archaea can assimilate short fragments of genetic material from viruses and plasmids and incorporate these sequences (known as protospacer) into the CRISPR locus. A crucial role to discriminate between self and non-self is attributed to Protospacer-associated motifs (PAM), which are short, conserved sequences adjacent to the protospacer and necessary for its incorporation. The protospacers are transcribed into crRNAs that, in case of a second infection, recognize its complementary target sequence within the invader's genome and redirect Cas nuclease activity to cut and disrupt it. Back in 2012, Jennifer Doudna and Emanuelle Charpentier envisioned to combine *Streptococcus pyogenes* Cas9 with an artificial RNA, that was called guide RNA (gRNA), to guide the nuclease to cleave a plasmid DNA or an oligonucleotide duplex bearing a protospacer sequence complementary to the gRNA and a bona fide PAM (Jinek *et al*, 2012). Cas9 belongs to type II family of CRISPR/Cas systems and can be directed in an RNA-dependent manner to a target sequence through a 20 bases gRNA. A 5'-NGG-3' consensus PAM sequence downstream the target sequence is required for Cas9 to open the double-strand DNA helix and for the formation of the RNA-DNA heteroduplex between the gRNA and its complementary sequence in the genome. The non-targeted strand forms a single-strand DNA 'R-loop' which is required for inducing conformational changes in Cas9 and activating its nuclease domains (Jinek *et al*, 2012; Jiang *et al*, 2016). Next, Cas9 triggers DNA DSBs thanks to its two catalytically active domains, HNH and RuvC, which cut the gRNA complementary and non-complementary strands, respectively, three bases upstream to the PAM sequence. Since its first application, the CRISPR/Cas9 platform has been employed to target the genome of different species and cell types, becoming by far the platform of choice for gene editing applications.

The requirement of NGG PAM downstream to the target sequence strongly constrains the range of sequences that can be targeted by Cas9. To overcome this limitation and broaden genome accessibility, Cas enzymes derived from other bacteria (Ran *et al*, 2015; Kim *et al*, 2017a; Müller *et al*, 2016; Hou *et al*, 2013; Esvelt *et al*, 2013) and evolved Cas9 variants with more relaxed PAM (Kleinstiver *et al*, 2015; Nishimasu *et al*, 2018; Hu *et al*, 2018; Edraki *et al*, 2019) can be also used. Researchers have recently developed a SpCas9 variant, known as SpRY, which necessitates a 5′-NRN-3' PAM sequence (Walton *et al*, 2020). This innovation has enabled the editing of genetic sites that were previously inaccessible, effectively addressing most of the limitations associated with PAM requirements.

3.4.2 Therapeutic applications of nuclease-based gene editing

Nuclease-mediated DNA DSBs trigger the activation of DNA damage response (DDR) pathways which can be exploited to disrupt a coding sequence or a regulatory element through NHEJ, or to correct a target sequence or integrate in situ a DNA sequence of interest through HDR (Ferrari *et al*, 2023). For the latter application, an exogenous DNA template is co-delivered alongside the gRNA and the Cas9. DNA templates for this purpose are single stranded oligodeoxynucleotide (ssODN) and non-integrating vectors, such as AAVs or integrase defective LVs (IDLVs).



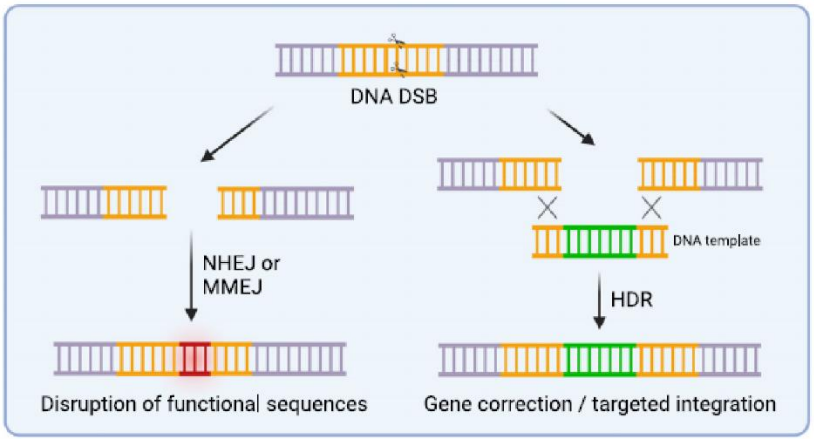


Figure 7. Schematic representation of Cas9-mediated DNA DSB and related repair outcomes. *Adapted from (Ferrari* et al*, 2023; Anzalone* et al*, 2020). Permission to reproduce from Cell Stem Cell under license number 5641461337114. Permission to reproduce from Nature Biotechnology under license number 5641470164164.*

3.4.2.1 Therapeutic editing through NHEJ

NHEJ-mediated gene editing can be exploited to achieve: (i) knock-out (KO) of a gene, (ii) targeted deletion or (iii) frameshift restoration:

(i) Optimized engineered nucleases can achieve in situ cleavage efficiency exceeding 90%. Consequently, when the nuclease activity is directed against a coding sequence, NHEJ induced insertion or deletions (indels) can KO a gene by altering its reading frame. Biallelic gene editing has the potential to yield a complete, enduring, and inheritable KO, making this approach valuable for both basic and translational research. In the last years, *CCR5* KO was proposed in human T cells and HSPCs to confer resistance to HIV infection (Tebas *et al*, 2014; Xu *et al*, 2019).

(ii) Two nucleases can be designed to simultaneously introduce DSBs upstream and downstream the target site, aiming to the deletion of the DNA sequence comprised between them. This approach may be relevant to increase KO efficiency, to delete one or more exons, to excide pathologically expanded nucleotide triplets, or to remove a regulatory region and tailor gene expression.

(iii) Indels induced by NHEJ may sometimes be used to rectify frameshift mutations, as in the context of some mutations causing Fanconi anemia (Román-Rodríguez *et al*, 2019).

3.4.2.2 Therapeutic editing through HDR

HDR-mediated gene editing can be exploited to achieve: (i) correction of point mutations, (ii) in situ gene correction, (iii) gene addition, (iv) transgene expression regulated by endogenous elements.

(i) In the context of genetic diseases caused by single-point mutation, a viable strategy involves guiding a nuclease near the mutation site and exploiting an HDR donor template carrying the wild-type sequence (Urnov *et al*, 2005). In contrast to GT approaches this method ensures the physiological regulation of

the gene. This approach can be applied in the context of sickle cells disease (SCD) caused by a point mutation in human β-globin gene (*HBB)* (Hoban *et al*, 2015; Dever *et al*, 2016; Romero *et al*, 2019) and of X-linked chronic granulomatous disease caused by mutations in *CYBB* (De Ravin *et al*, 2017).

(ii) Since most genetic diseases are caused by different mutations within the same gene, an alternative strategy is to deliver an HDR donor template comprising the cDNA for the disease-causing gene. This approach enables functional correction of nearly all potential gene mutations with a single strategy ("one-size-fits-most"), ensuring that gene expression remains under endogenous regulation. The optimization of the codon usage or the incorporation of introns further helps in achieving precise physiological expression. In situ gene correction has been investigated for different hematological diseases, such as β-Thalassemia (Voit *et al*, 2014), SCID-X1 (Schiroli *et al*, 2017), Hyper-IgM 1 syndrome (Kuo *et al*, 2018; Vavassori *et al*, 2021) and WAS (Rai *et al*, 2020).

(iii) The installment of a therapeutic cassette within a specific genomic region can be a valuable approach when precise physiological regulation is not critical, or when achieving gene overexpression is necessary for therapeutic efficacy (Moehle *et al*, 2007). Targeted integration can be direct to a 'safe harbor' region of the genome, where cassette integration does not disrupt the expression of nearby genes, ensures stable gene expression across various cell types, prevents cassette silencing and minimize the risk of insertional mutagenesis (Lombardo *et al*, 2011; Sadelain *et al*, 2011). Two loci in the human genome have been often exploited as safe harbors, the Adeno-associated virus site 1 (*AAVS1*) and the *CCR5* gene. *AAVS1* is the preferred integration site for adeno-associated viruses serotype 2 and is located on chromosome 19, specifically between exon 1 and intron 1 of the protein phosphatase 1 regulatory subunit 12C (*PPP1R12C*) gene (DeKelver *et al*, 2010).

(iv) Leveraging on endogenous regulatory elements for the control of transgene expression can yield highly robust and cell-specific protein expression. For example, transgene can be inserted under the control of α-globin promoter to achieve sustained expression within the erythroid lineage (Pavani *et al*, 2020)

3.4.3 Hurdles of nuclease-based gene editing in HSPCs

Although gene editing has proven to be a versatile platform with successful applications in various organisms and cell types for different purposes, its application in human HSPCs still faces limitations, particularly in terms of efficiency when aiming to HDR editing, and safety due to potential genotoxicity of inadvertent DNA repair outcomes and detrimental cellular responses triggered by cell and genome manipulation.

3.4.3.1 Delivery systems for nuclease-based gene editing in HSPCs

The choice of the systems for the delivery of programmable nucleases and the DNA donor template, when aiming to HDR, is one of the major determinants of HSPC editing efficiency. Regarding the nucleases, they can be either delivered as *in vitro* transcribed mRNA (Genovese *et al*, 2014; Schiroli *et al*, 2017) or as ribonucleoprotein (RNP) complex of Cas9 protein pre-assembled with the gRNA (Hendel *et al*, 2015; Schiroli *et al*, 2019a). These methods ensure a high but transient expression of the nuclease, thus limiting the possible off-target effects, particularly for RNP (Dever *et al*, 2016). Among the route of administration, electroporation is the gold standard in the field to deliver nucleases and gRNAs (Genovese *et al*, 2014), however this method induces a basal level toxicity (Batista Napotnik *et al*, 2021) and detrimental transcriptional responses that reduce clonogenic capacity and repopulation potential of HSPCs. Recently, lipid nanoparticles (LNPs) have been used as delivery methods of the editing components resulting in improved cellular viability without reducing editing efficiency (Vavassori *et al*, 2023). Concerning the donor template, in a seminal paper viral delivery of both nuclease and DNA donor template resulted in low editing efficiency in HSPCs (Lombardo *et al*, 2007). Conversely, delivery of the nuclease by electroporation and of the DNA donor template by viral transduction significantly improved HDR editing efficiency (Genovese *et al*, 2014). Different viral templates can be employed for HDR editing in HSPCs. IDLVs were initially chosen as delivery vehicle as they were found more efficient and less toxic when compared to plasmids in human HSPCs (Genovese *et al*, 2014). Beyond IDLVs, other viral vectors have been exploited as DNA delivery methods. In particular, the combination of AAV serotype 6 (AAV6) transduction and nuclease electroporation resulted in successful editing of HSPCs, achieving significant levels of long-term targeted integration, typically in the range of 10% to 20%, upon in vivo transplantation (Dever et al., 2016; Wang et al., 2015). Another suitable vehicle for the

HDR template is ssODN, which can be co-delivered together with the nuclease by electroporation. The three primary delivery vehicles for therapeutic gene editing via HDR, (IDLVs, AAV6 and ssODNs), exhibit distinct characteristics in terms of efficiency in both *in vitro* and *in vivo* settings, cargo capacity, and toxicity levels (Romero *et al*, 2019; Allen *et al*, 2023). On one hand, AAV6 vectors demonstrate superior efficiency both *in vitro* and *in vivo* when compared to IDLVs. However, AAV6 has a more limited cargo capacity, accommodating up to 4.8 kb, compared to IDLVs (up to 7 kb of genetic material). On the other hand, ssODNs exhibit efficiency levels similar to AAV6 but have mostly limited cargo capacity. Hence, the choice of the delivery vehicle depends on the specific requirements of the gene editing experiment, considering factors such as cargo size, efficiency, versatility, and toxicity, to achieve the desired outcomes effectively and safely (Dever & Porteus, 2017).

3.4.3.2 Editing efficiency in human HSPCs

The development and optimization of nucleases as well as the engineering of stable gRNA and the refinement of their delivery methods paved the way for remarkably efficient NHEJ-mediated gene disruption in human HSPCs, reaching successful results in the clinics (Wu *et al*, 2019; Fu *et al*, 2022; Frangoul *et al*, 2021). In contrast the low proficiency of HDR in LT-HSCs is a critical limit of these gene editing approaches. LT-HSCs exhibits poor transduction efficiency and limited repair through HDR, likely due to slow cell cycling and low expression levels of DNA repair machinery within this cell population. A crucial role in determining editing efficiency by HDR is assigned to the *ex vivo* culture condition, which must strike a balance between preserving the stem cell properties and promoting proliferation. To balance preservation of stem cell and cell cycle activation, in gene editing applications HSPCs are pre-stimulated for 48-72 hours after thawing before undergoing editing procedure. Therefore, on one hand shortening time in culture is essential to better preserve HSPC repopulation potential; on the other hand, cell culture and cycling enables HSPCs to enter S/G2 phases of the cell cycle and favor DNA DSB repair by HDR in presence of the donor templates. State-of-the-art culture conditions for HSPCs include several cytokines, i.e., stem cell factor (SCF), thrombopoietin (TPO), FMS-related tyrosine kinase 3 ligand (FLT-3L). This cytokine cocktail promotes HSPC activation and survival *in vitro* and support preservation of the stem cell properties to some extent (Walasek *et al*, 2012). StemRegenin 1 (SR1), a purine

residue antagonist of the aryl hydrocarbon receptor, is often included in the HSPC culture media as it can increase HSPC engraftment upon xenotransplant in NSG mice (Walasek *et al*, 2012; Boitano *et al*, 2010). Furthermore, a chemical library screening identified UM171 that in combination with SR1 showed greater HSPC expansion (Fares *et al*, 2014). Recently, the addition of human interleukins (IL) in the HSPC culture medium was proposed to balance expansion and stemness maintenance (Rai *et al*, 2023). While IL3 was reported to support progenitor expansion and promote HDR, IL6 preserved HSC at the cost of reduced expansion and HDR efficiency. Further HSPC expansion can be achieved by using 3D culture systems resembling the BM niche as zwitterionic hydrogel which leads around 70-fold expansion assessed by limiting dilution assay and in xenotransplant experiments (Bai *et al*, 2019a). In the last months, Yamazaki and colleagues developed a cytokine-free medium in which only UM171 and a phosphoinositide 3-kinase activator support the *ex vivo* expansion of CB-derived HSPCs (Sakurai *et al*, 2023). These strategies have not been explored in the context of gene editing yet.

To enhance HDR efficiency many strategies have been tested mostly aiming to inhibit NHEJ (Maruyama *et al*, 2015; Chu *et al*, 2015) or to favor HDR by promoting cell cycle entry (Charpentier *et al*, 2018; Jayavaradhan *et al*, 2019; Ferrari *et al*, 2020a; Shin *et al*, 2020). Currently, HDR efficiency remains a limiting factor for some gene editing applications especially when edited cells, or their progeny, do not bear a selective advantage over the non-edited counterpart.

3.4.3.3 Cellular responses limiting efficiency and tolerability of HSPC gene editing

The gene editing process has the potential to impact the functionality of HSPCs and to induce toxicity. Notably, AAV transduction, exogenous nucleic acids, and the generation of DNA DSBs may affect HSPCs. It has also been observed that genome editing treatments inherently exhibit higher toxicity when applied to primary cells in comparison to cell lines, with HSPCs being particularly sensitive to gene targeting procedures (Urnov *et al*, 2005). Throughout evolution, human cells have developed defense mechanisms against bacterial and viral infections. Contaminants present in vector preparations have the potential to trigger innate cellular responses through the activation of pattern recognition receptors (PRRs) following electroporation procedures (Sadler & Williams, 2008). To mitigate this response, cesium-chloride purification methods are available to

eliminate empty particles from AAV vector preparations, while chromatographic purification techniques can enhance the purity of IDLVs. Furthermore, foreign RNAs present in the cytoplasm, such as mRNA encoding nucleases, have the potential to activate innate cellular responses and trigger the production of Type I interferons (IFN). This IFN response, in turn, initiates a series of changes in HSPCs, including reprogramming and exit from the G0 phase, leading to differentiation, proliferation, and apoptosis (Essers *et al*, 2009; Passegué & Ernst, 2009; Liu *et al*, 2012). To mitigate these innate cellular responses, chemically modified mRNAs (Warren *et al*, 2010) and gRNAs can be exploited (Hendel *et al*, 2015). Furthermore, a series of purification steps including high performance liquid chromatography (HPLC), are essential to eliminate contaminants in the Cas9 preparation and reduce cellular toxicity (Rajagopalan *et al*, 2018; Karikó *et al*, 2011). Despite high-quality reagents are crucial in mitigating the IFN response, the intrinsic toxicity associated with gene editing procedures cannot be fully eliminated. HSPCs are indeed susceptible to the activity of nucleases, as they rapidly activate the DDR when a DNA DSBs is introduced into the genome, which can potentially affect HSPC functions (Milyavsky *et al*, 2010). Moreover, the nuclease off-target activity may induce additional DNA DSB increasing the DDR response and leading to differentiation, replication arrest, apoptosis, or senescence. Ultimately, this can reduce the long-term engraftment capacity of HSPCs. Importantly, this DDR response is observed even in the most primitive HSPC compartment, where p53-dependent DDR activation is the predominant and almost exclusive response to even one DNA DSB (Schiroli *et al*, 2019a). Activation of the p53 pathway induces a temporary slowdown in cell cycle progression, with no significant impact on HSPC repopulation potential. However, when multiple DSBs are induced by a low-specificity nuclease, they lead to a robust p53 response, including the activation of inflammatory transcriptional programs resulting in prolonged cell cycle arrest and a significant negative impact on repopulation capacity. Crucially, the p53 response is worsened even when highly specific nucleases are combined with AAV6 transduction, causing a response comparable to that observed with low-specificity nucleases. This highlights the pivotal role of the AAV vector in triggering a cumulative DDR response, which ultimately shrink the human graft size and the oligoclonal reconstitution (Ferrari *et al*, 2020a). The induced cell cycle arrest and proliferation delay can be partially alleviated by transient p53 inhibition, achieved by co-electroporating an

mRNA encoding for GSE56, a dominant-negative of p53. GSE56 reduces the impact on cell cycle progression and increases the extent of *in vivo* repopulation by edited HSPCs (Schiroli *et al*, 2019a). Furthermore, the GSE56 mediated p53 inhibition can be combined with the delivery of the adenoviral protein E4orf6/7 (GSE56/E4orf6/7) which push cell cycle progression resulting in improved HDR editing without perturbation of HSPCs' repopulation potential (Ferrari *et al*, 2020a).

3.4.3.4 Off-target editing and on-target genotoxic byproducts of DNA DSB repair

Genotoxicity is one of the major concern of gene editing applications. Indeed, the introduction of DNA DSB not only activate DNA damage response but may also result in genotoxic events at the target locus, such as large deletions, translocations, and chromosomal rearrangements (Leibowitz *et al*, 2021; Adikusuma *et al*, 2018; Park *et al*, 2022; Turchiano *et al*, 2021; Kosicki *et al*, 2018). Notably, recent research has highlighted the pivotal role of the delivery template choice in influencing genotoxicity at both on- and off-target sites. Therefore, in the case of AAV6, its inverted terminal repeats can become integrated into the genome during the editing process, activating transcription of nearby genes at both on- and off-target sites. On the other hand, IDLV delivery appears to mitigate this effect (Ferrari *et al*, 2022b). Therefore, nuclease-based gene editing results in a genetically heterogenous populations of treated cells within the final product. A large fraction of them lacks therapeutic efficacy, and rather bears an undetermined genotoxic risk, which is likely context-specific and remains problematic to assess. Another source of genotoxicity that must be particularly considered when aiming for clinical translation, arises from off-target effects. Sequences in the genome containing stretches of bases highly homologous to the intended gRNA target sequence can undergo DNA DSB. In case an off-target site is located within a gene or a regulatory region, it may result in the loss of its normal function. Even if the DNA DSB occurs within a neutral, dispensable region, it could still generate chromosomal rearrangements at the off-target site or with the on-target site, and chromosomal loss (Guo *et al*, 2023).

To mitigate Cas9 off-target activity, several strategies have been devised. mRNA or RNP delivery of Cas9 reduces off-target activity compared to plasmid or viral delivery, likely due to faster degradation, and thus shorter half-life, of the nuclease. Fu and colleagues introduced a modified gRNA, referred to as truncated guide RNA (trugRNA), which features a shorter complementary region of 17-18 base pairs (bp) instead of the

typical 20. Remarkably, this trugRNA demonstrated no differences in on-target activity while exhibiting a remarkable over 5,000-fold reduction in off-target mutagenesis (Fu *et al*, 2014). Another avenue of improvement focuses on the Cas9 protein itself. One approach involves the strategic inactivation of key domains, such as HNH or RuvC, within Cas9, yielding a nickase Cas9 (nCas9) variant that introduces single-strand breaks (SSBs) rather than DSBs. By employing a paired-nickases strategy, it is possible to enhance on-target activity and generate 5' overhang ends (Ran *et al*, 2013). Similarly, researchers have explored the fusion of a dead Cas9 (dCas9) with a FokI catalytic domain as a potential option to reduce off-target effects (Tsai *et al*, 2014). This modification ensures that FokI cleaves only when dimeric RNA-guided FokI-dCas9 nucleases are present. Additionally, an enhanced specificity SpCas9 (eSpCas9) variant has been developed through structure-guided mutagenesis. This variant incorporates three substitutions that reduce its positive charge, thereby decreasing its interaction with the non-target DNA strand (Slaymaker *et al*, 2016). Furthermore, a high-fidelity version of SpCas9, known as SpCas9-HF1, has been engineered by eliminating specific hydrogen bonds that facilitate gRNA-target DNA binding. This alteration promotes high-affinity interactions and reduces the occurrence of off-target activity (Kleinstiver *et al*, 2016).

3.5 Nickase-based gene editing

In the last years, the gene editing toolbox have been expanded thanks to the advent of base and prime editing systems, which hold the promise for precise genetic modification without requiring DNA DSB formation (Anzalone *et al*, 2020). At variance with CRISPR/Cas nucleases, base editors (BEs) and prime editors (PEs) are chimeric molecules currently capable of installing mostly short-range modifications by tethering deaminase or reverse transcriptase activity, respectively, to the target site and simultaneously delivering DNA SSB to favor fixation of the intended edit.

3.5.1 Base editing

BEs, which have been developed in 2016 from David Liu's laboratory, have undergone continuous evolution since their establishment (Komor *et al*, 2016). This innovative platform can generate specific transition point mutations when used in combination with a gRNA. The power of this editing system relies on its ability to theoretically correct a sizable fraction of point mutations causing human diseases (Rees & Liu, 2018b) without

the requirement of DNA DSBs. Notably, BEs consist of two key components: i) a deamination domain that directly modifies nucleotides within a defined editing window on one of the two genomic strands, and ii) a nCas9 that induces a SSB on the opposite strand to facilitate more effective editing. Depending on the type of transition point mutation, base editors are classified in Cytosine (C-) BE (C•G transition to T•A) and Adenine (A-) BE (A•T transition to G•C).

3.5.1.1 Cytosine BE

In the first generation of BEs, known as BE1, Komor and colleagues combined catalytically inactive Cas9, dCas9 with a cytidine deaminase domain APOBEC1 through a 16 amino acids (aa) linker (Komor *et al*, 2016). Although dCas9 retains its ability to bind to DNA in an RNA-guided manner, it lacks the capacity to cleave the DNA because of the Asp10Ala (D10A) and His840Ala (H840A) mutations inactivating its two catalytic domains (Qi *et al*, 2013) . Consequently, when combined with a gRNA, dCas9 localizes to the target location and generates a R-loop complex in which the genomic strands are displaced (Jiang & Doudna, 2017). This single-strand exposure enables the cytidine deaminase to catalyze C deamination into uracil (U), which mimics the base-pairing properties of thymine (T). Deamination occurs in a specific location corresponding to protospacer position 4-8, defined as editing window. However, the presence of a U•G mismatch in the human genome can activate BER, thereby limiting BE1 efficiency, which only reached 2-8% in cell lines (Lee & Kang, 2019a). Given that C deamination is a common event in eukaryotic cells, with a rate of 100-500 spontaneous cytosine deamination per cell per day (Lindahl, 1993), U bases in the genome are promptly identified and eliminated by uracil glycosylase (UG), the apical enzyme of the BER pathway. UG creates an abasic site and subsequently triggers APEX1 endonucleases activation, leading to nick formation. The nick is then sealed using the complementary strand as a template and restoring the cytosine. Therefore, in the second generation of the editor, known as BE2, a UG inhibitor (UGI) was fused to the dCas9 domain through a 4 aa linker resulting in improved editing efficiency reaching up to 20%. Nevertheless, even with BER inhibition, the U•G mismatch induced by BE2 can still be corrected using either the edited or the non-edited strand as a template, thereby reducing the chance of fixing the intended edit rather than restoring the native base. To push the installment of the desired modification by employing the edited strand as a template for U•G mismatch

resolution, Komor and colleagues developed the third generation of the editor, BE3. In this version, the dCas9 was replaced with a D10A nCas9, introducing a nick in the non-edited strand. Consequently, the non-edited strand containing the G mimics a newly synthesized strand and is preferentially resolved by MMR into the desired U:A and T:A products. BE3 achieved an efficiency of approximately 35%, with a limited fraction of indels ranging from 0.5% to 2%, at the target site. These indel levels were higher when compared to the BE1 and BE2 versions, possibly ascribed to SSB conversion into DSB during cell replication. Fourth generation BE, BE4, was further optimized by (i) extending cytidine deaminase-nCas9 linker from 16 to 32 aa, (ii) extending nCas9-UGI linker from 4 to 9 aa and (iii) adding a second UGI domain and resulted in 1.5-fold improved efficiency and 2-fold reduction in indels formation compared to the BE3 generation (Komor *et al*, 2017). A further step in the evolution of CBE was the generation of BE4max which contains bipartite nuclear localization signal at N- and C-termini and optimized codon usage and resulted in 3-fold prolonged expression and 1.7-fold improved efficiency compared to the original BE4 (Koblan *et al*, 2018a). Beside APOBEC1, different cytidine deaminase domains can be fused with nCas9 for various purposes. Mutated APOBEC1 domains enable a narrower editing window (Kim *et al*, 2017c), whereas APOBEC3 variants and AID variants expand the editing window. Furthermore, different deaminases exhibit preferential activity on different substrates, such as the TC motif for APOBEC1 and CC for APOBEC3 (Anzalone *et al*, 2020).

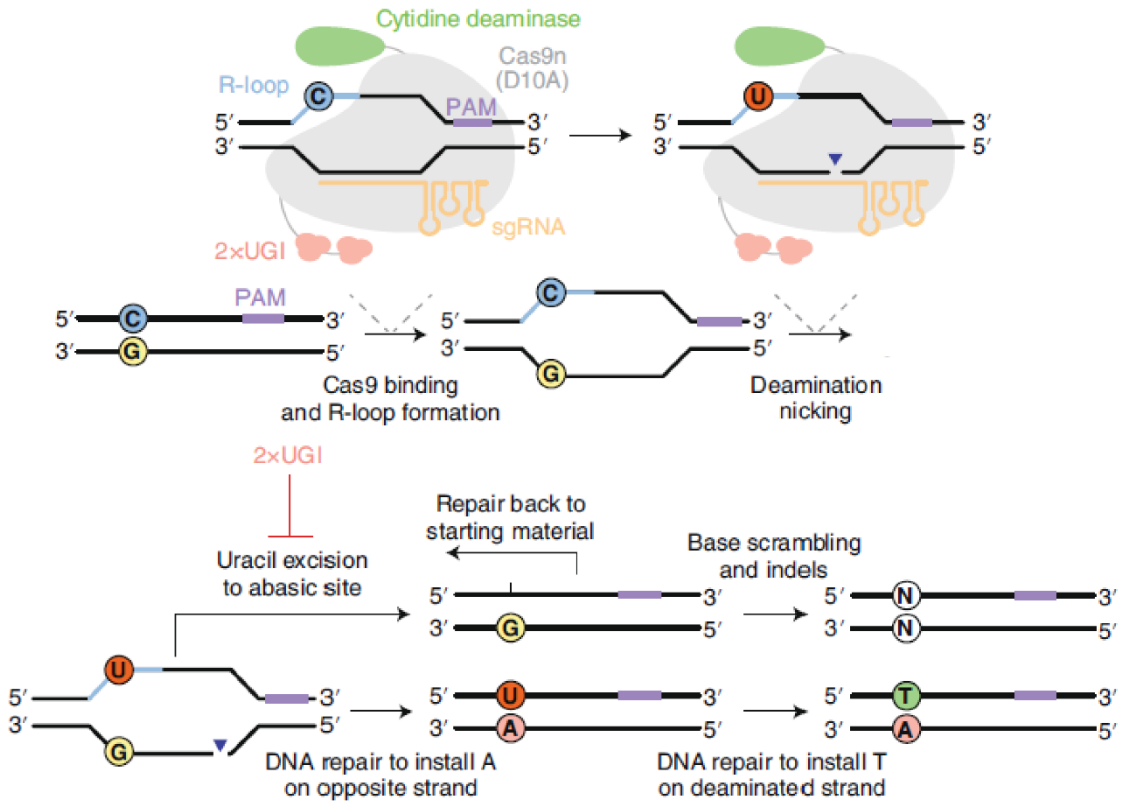


Figure 8. Schematic representation of CBE and its related repair outcomes. *Adapted from (Anzalone* et al*, 2020). Permission to reproduce from Nature Biotechnology under license number 5641470164164.*

3.5.1.2 Adenine BE

As discussed earlier, eukaryotic cells are daily exposed to spontaneous cytosine deamination that, if left unrepaired, result in C•G to T•A mutations. Therefore, from an evolutionary standpoint, human genomes tend to accumulate T•A mutations, which account for approximately 45% of the pathogenic variants in the ClinVar database (Rees & Liu, 2018b). Consequently, ABEs, which convert A into G, hold promise for the correction of a large number genetic disease. Although A can be deaminated into inosine (I), which exhibits G preferential base-pairing, a key challenge in developing ABEs was the absence of known adenosine deaminase enzymes acting on DNA. In a pioneering effort, Gaudelli and colleagues fused a naturally occurring adenine deaminase from *Escherichia coli*, TadA, with nCas9 D10A (Gaudelli *et al*, 2017). However, no A-to-G modifications were observed when this construct was combined with a gRNA and delivered in HEK293T cells. Consequently, they devised a bacterial selection method and evolved TadA to act on DNA, leading to the creation of the TadA variant (TadA*). Despite

its effective A-to-I conversion in bacterial cells, the TadA*-nCas9 fusion protein demonstrated limited efficiency in human cells. This disparity can be attributed to the fact that the TadA enzyme operates as a homodimer in bacterial cells. While during selection in bacterial cells, TadA* and endogenous TadA can form heterodimers, the absence of TadA in human cells prevents the formation of the dimer. To address this issue, a TadA-TadA*-nCas9 polypeptide was developed, resulting in significantly enhanced editing efficiency, averaging 20% in human cell lines. Subsequent rounds of selection, with increasing selective pressure, enabled the achievement of approximately 60% editing efficiency using ABE7.10. As for the development of BE4max, Koblan and colleagues introduced bipartite nuclear localization signals at both the N- and C-termini, along with optimized codon usage for ABE7.10 (Koblan *et al*, 2018a). This modification resulted in the creation of ABEmax, which exhibited a 1.5-fold improvement in editing efficiency. More recently, other ABE versions, such as ABE8e (Richter *et al*, 2020) and ABE8.20-m (Gaudelli *et al*, 2020a), were generated using phage-assisted evolution and directed evolution techniques, respectively. These versions of the editor have demonstrated superior efficiency, reaching up to 90% in human primary cells, becoming the most widely used ABE tools.

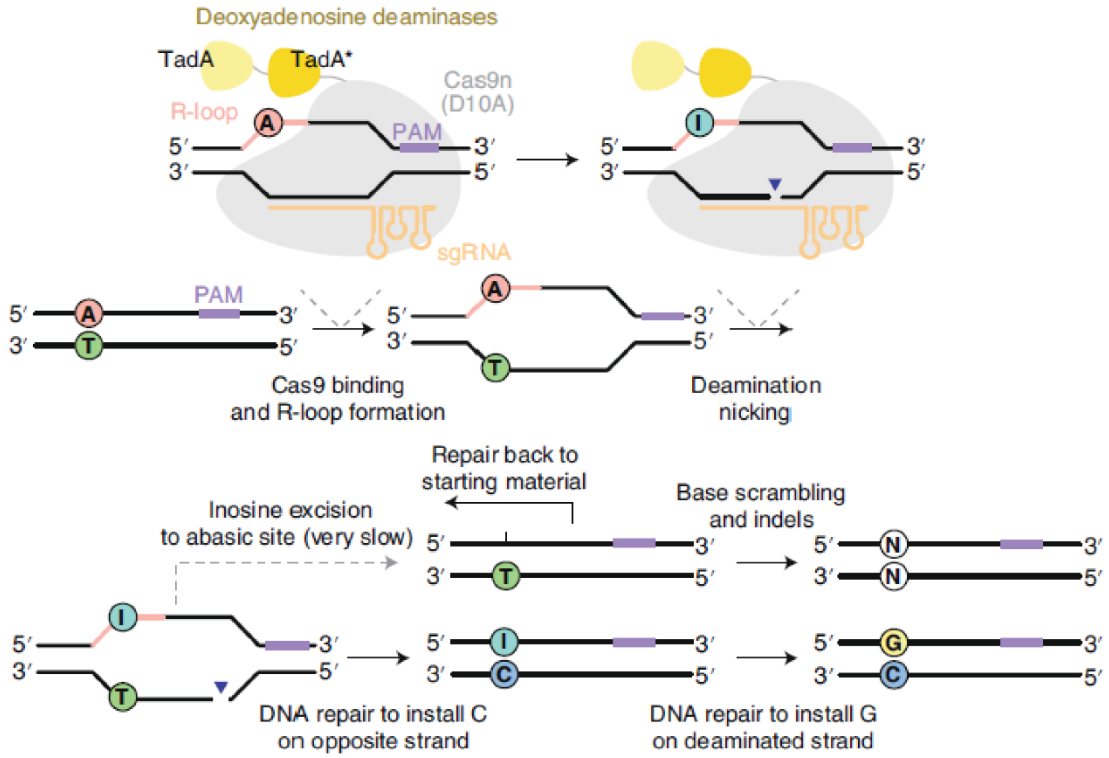


Figure 9. Schematic representation of ABE and its related repair outcomes. *Adapted from (Anzalone* et al*, 2020). Permission to reproduce from Nature Biotechnology under license number 5641470164164.*

3.5.1.3 Base editing applications in human HSPCs

In 2020, Zeng and colleagues developed a therapeutic approach using CBE to correct SCD and β-thalassemia (β-thal) patients derived HSPCs (Zeng *et al*, 2020b). Specifically, they combined CBE protein with a gRNA targeting *BCL11A* erythroid enhancer and delivered the RNP complex in patients' cell to reactivate fetal hemoglobin (HbF). In SCD and β-thal HSPCs they reached 70% editing efficiency with one cycle of electroporation and 90% efficiency with two cycles of electroporation, albeit at the cost of a substantial reduction in viability (83% to 47%). Moreover, CBE edited cells from heathy donor persisted long term upon serial xenotransplantation, despite the percentage of edited cells tented to decrease compared to the input HSPCs. Regarding ABE, different strategies were successfully tested in the context of hemoglobinopathies as SCD and β-thal aiming to reactivate fetal hemoglobin (HbF) expression which compensate for the deficit of normal hemoglobin, or to specifically correct disease-causing mutations. Regarding HbF induction, Gaudelli and colleagues demonstrated a 60% editing efficiency in HSPCs by introducing the 'British mutation,' a non-pathogenic variant associated with HbF persistence in adults (Gaudelli *et al*, 2020a). Antoniou and colleagues achieved HbF expression by delivering ABE mRNA to induce mutations approximately 200 bp upstream of the γ-globin (*HBG*) transcriptional start site in patients' HSPCs (Antoniou *et al*, 2022). More recently, ABE protein was delivered in β-thal patients' HSPCs, targeting the *BCL11A* enhancer or *HBG* promoter, resulting in over 80% efficiency and HbF reactivation (Liao *et al*, 2023). Additionally, ABE was used to introduce a mutation into the *HBG* promoter at position -175, leading to clinically relevant levels of editing in xenotransplant experiments accompanied by HbF expression (Mayuranathan *et al*, 2023). Regarding the correction of disease-causing mutation, in 2021, Newby and colleagues exploited *ex vivo* delivery of ABE mRNA to convert the SCD mutation (A>T) into a non-pathogenic variant (Newby *et al*, 2021). They achieved 80% efficiency in patients' cells, which was maintained at 70% efficiency in the long-term engraftment of immunodeficient mice. Furthermore, ABE was combined with Cas9 variants with relaxed PAM recognition and used to correct one of the most frequent β-thal mutation (IVS1-110)

in patient-derived HSPCs, achieving 80% correction (Hardouin *et al*, 2023). Notably, ABEs are now approaching the clinical arena for the treatment of hemoglobinopathies (NCT05456880). Beside hemoglobinopathies, ABE was used to correct the most frequent mutation causing Fanconi Anemia, reaching from 40 to 60% efficiency depending on the HSPC donor (Siegner *et al*, 2022) and the mutation causing CD3δ SCID reaching 88% efficiency in xenotransplant experiments (McAuley *et al*, 2023).

3.5.1.4 Base editing limitations and improvements

Despite the significant advancements provided by base editing in the gene editing field, some challenges remain to be considered: (i) indels and unexpected point mutation, (ii) target sequence constrains, (iii) editing window, PAM constrains and bystander editing, (iv) off-target activity.

(i) Although base editing is reasonably less genotoxic and more precise compared to a Cas9 editing approach because of the lack of DNA DSB requirement, different undesired events may arise at the target site. Concerns include conversion of the DNA SSB into DSB during cell replication or in response to cellular repair mechanisms, such as BER, as well as unintended mutations due to activation of error prone repair mechanism, such as MMR (Anzalone *et al*, 2020).

(ii) While CBE and ABE are effective at converting C•G to T•A or A•T to G•C bp, the conversion of other types of bp is more challenging and less efficient thus limiting the application in some context disease. Notably, recent advancements have led to the development of evolved BE variants designed to address this issue. In particular, BEs were developed to induce C•G transversions while minimizing C•T or C•A mutations (Kurt *et al*, 2020; Zhao *et al*, 2021; Koblan *et al*, 2021; Chen *et al*, 2021a). However, their application is currently limited to cell lines, and, up to now, not all possible base pairs are editable by BEs.

(iii) Deaminase activity occurs in a specific range of bases of the protospacer region, known as editing window, which becomes exposed during formation of the R-loop. This range usually spans from positions 2 to 8 distal from the PAM sequence. The availability of a suitable NGG-PAM sequence plays a pivotal role in determining whether a point mutation can be efficiently

corrected by BEs. As discussed for Cas9 nuclease this limitation can be overcome by fusing deamination domain with Cas orthologues (Ran *et al*, 2015; Kim *et al*, 2017a; Hou *et al*, 2013; Esvelt *et al*, 2013)or with evolved Cas9 thus broadening the genome accessibility (Kleinstiver *et al*, 2016; Nishimasu *et al*, 2018; Hu *et al*, 2018; Edraki *et al*, 2019; Walton *et al*, 2020). Furthermore, the sequence composition of the DNA flanking the desired base requires careful consideration when aiming to correct disease-causing point mutations. If other bases amenable to editing exist within the editing window, they can also undergo modification, a phenomenon termed 'bystander editing' resulting in mutagenic events at the target locus. To mitigate bystander editing, researchers developed CBE and ABE variants bearing different editing windows (Kim *et al*, 2017b). Apart from the sequence limitations imposed by the PAM and activity window requirements, the specific variant of the deaminase enzyme embedded in the BE has its own sequence context preferences influencing editing efficiency (Komor *et al*, 2016; Wang *et al*, 2018; Gehrke *et al*, 2018).

(iv) As for nucleases, off-target point mutations and indels induced by BEs may mutate coding sequences or regulatory regions, potentially affecting its physiological activity. Importantly, the life-long nature of HSPC treatments imposes a careful evaluation of such events. Indeed, for each possible application off-target risk should be carefully evaluated and weighed against the potential benefit of the treatment. Two types of off-target activity exist for BEs. (i) The nCas9 activity due to sequence homology of the gRNA with other sequences in the genome may lead to gRNA-dependent off-target activity, similarly to what happens with nucleases. Although the outcome of gRNA-dependent off-target activity could be often more favorable compared to a nuclease Cas9 programmed with the same gRNA due to the less disruptive activity of the former, gRNA dependent off-targets must be still carefully evaluated when aiming to translate a base editing strategy into clinical practice. As for Cas9 nuclease, variant with increase specificity such as high fidelity Cas9 can be employed to reduce 'gRNA-dependent off-target' (Hong & He, 2023). (ii) The transient overexpression of catalytically active deamination

domains exposes the cells to an intrinsic risk of genome- and transcriptome-wide DNA or RNA gRNA-independent off-target activity. Indeed, unintended deamination events may occur at transcription-associated R-loops. While RNA gRNA-independent off-target have been predominantly attributed to ABE (Gaudelli *et al*, 2020a; Grünewald *et al*, 2019), DNA gRNA-independent off-target is mostly triggered by CBE (Jin *et al*, 2019; McGrath *et al*, 2019; Doman *et al*, 2020). These events, despite difficult to be evaluated because of their reasonably low and stochastic nature, must be taken into account since they can lead to mutational overload, exposing cells to functional alterations and potentially oncogenic hits. Furthermore, C>T mutation are the most common across different human cancer thus further imposing a careful evaluation of such events (Kandoth *et al*, 2013; Welch *et al*, 2012).

3.5.2 Prime editing

Another promising cutting-edge gene editing tool is prime editing, which is capable of generating all the types of single nucleotide variants (SNVs) and intended small indels (Anzalone *et al*, 2019b). Prime editors are chimeric protein composed of a nCas9 enzyme (with inactivated HNH catalytic domain) fused with a Reverse Transcriptase (RT) domain sourced from the Moloney Murine Leukemia Virus. PE is paired with a prime editor gRNA (pegRNA), which serves to initiate a nCas9-mediated DNA SSB, while simultaneously providing the template for reverse transcription originating from the cleaved strand. First, upon binding to the target site the Cas9 RuvC nuclease domain creates a nick in the DNA strand that contains the PAM sequence. Second, the primer binding sequence (PBS), which is contained in the 3' region of the pegRNA, binds the 3' free ends in the genomic DNA. Third, the genomic 3' free end is utilized as a primer for reverse transcription using the RT template which includes the desired edits along with a region of homology to the target site. Upon reverse transcription, a 3' flap containing the newly synthesized edited DNA is formed, which compete for incorporation with the unedited 5' flap. Since 5' flaps are naturally recognized from 5' endonucleases involved in DNA repair mechanisms, the edited 3' flap can be incorporated. This process results in a heteroduplex DNA formation comprising one edited and one non-edited strand. Consequently, the permanent integration of the edit is achieved by MMR pathway only if the edited strand is used as template for heteroduplex resolution. In the PE3 version of

prime editing, the latter process is facilitated by using a gRNA to guide PE2 in nicking the non-edited strand and to stimulate its resynthesis. A further implementation of the PE3 system, known as PE3b, employs a specialized nicking gRNA that selectively targets the edited sequence. This strategic approach aims to delay nicking of the non-edited DNA strand until the conversion of the other strand into the edited sequence has been completed, thus reducing the generation of indel byproducts. More recently, improved versions of PE were evolved upon the identification of DNA repair mechanisms activated during prime editing that may hamper efficient installment of the prime edit. A CRISPR interference-based screening revealed that the MMR pathway strongly counteracts heteroduplex resolution (Chen *et al*, 2021b). Transient inhibition of MMR, accomplished through the use of the dominant-negative protein MLH1dn, resulted in a remarkable up to 7-fold enhancement in prime editing efficiency, observed both in cell lines and primary T cells. MLH1dn was integrated into the 3' region of PE2 and PE3, leading to the development of PE4 and PE5, respectively. Furthermore, the PE architecture was improved through the following strategies: (i) optimizing the codon usage of the RT domain, (ii) introducing a 34-nt linker between nCas9 and the RT domain, incorporating two nucleotide localization signals, (iii) integrating R221K and N394K mutations into SpCas9, enhancing its nuclease activity, and (iv) including an additional NLS domain at the C-terminal. This resulting PE construct was termed PEmax. Liu's group has recently developed PE6 through phage-assisted evolution of RT domain (Doman *et al*, 2023). This engineering advancement enables a reduction in the size of the construct of approximately 600 bases, thereby simplifying its delivery. Furthermore, they generated different PE6 variants (a-g) by utilizing RT specialized in different types of modifications, resulting in a successful enhancement of editing efficiency in cell lines and primary T cells.

In addition to single-point mutations and small indels, PEs have also been employed in the generation of larger genomic modifications. Firstly, the combination of dual PEs, each associated with their respective pegRNA targeting the two opposite strands, has enabled the deletion and replacement of regions ranging from 1 to 10 kb (Jiang *et al*, 2022). An analogue approach has also been used to invert a target region spanning 40 kb (Anzalone *et al*, 2022). Secondly, Wang and colleagues developed a system in which the two pegRNAs contain a homologous region at the PBS in the 3' end, giving rise to reverse transcription in both genomic orientations (Wang *et al*, 2022). This design prevents the

formation of a 3' flap and allows for the insertion of sequences up to 1 kilobase. However, it's worth to note that the efficiency of this method inversely correlates with the size of the insertion. Finally, the PASTE editor was developed which consist of a nCas9 fused with both RT and serine integrase and allowed to reach targeted insertion up to 37kb (Yarnall *et al*, 2023). Although these systems hold the promise for targeted integration approaches that avoid the recruitment of HDR machinery and the formation of DNA DSB, their efficiency remains limited. Additionally, the complexity of these methods may pose challenges when applied to primary cells.

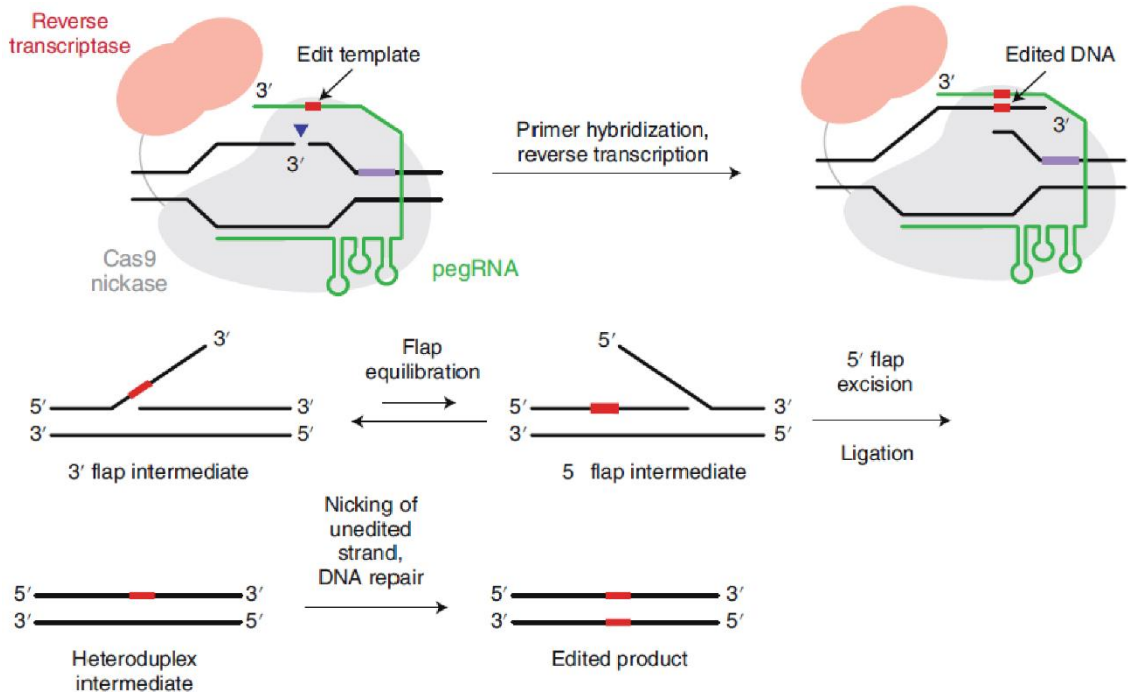


Figure* 10. *Schematic representation of PE and its related repair outcomes. *Adapted from (Anzalone* et al*, 2020). Permission to reproduce from Nature Biotechnology under license number 5641470164164.*

3.5.2.1 Prime editing applications in human HSPCs

Prime editing efficiency currently remains modest compared to other editing platforms, often limiting its application in human primary cells. Particularly, PE3 has been tested in human primary T cells however reaching up to 8% efficiency with half alleles carrying imprecise modifications and/or byproduct (Petri *et al*, 2022a). Higher efficiencies have been achieved in human primary T cells by using the PE4 and PE5 editors, ranging from 30% to 60%, depending on the target locus (Chen *et al*, 2021c). Recently, prime editing has been applied to human HSPCs. In a first report from Liu's group, they demonstrated an efficiency ranging from 10% to 80% in healthy HSPCs,

depending on the target locus, using optimized PEmax reagents (Everette *et al*, 2023). Furthermore, they reported an efficiency of 20-40% in HSPCs derived from SCD patients from four different donors, with an indel frequency of nearly 5%. Importantly, the editing efficiency was maintained after xenotransplantation into immunodeficient mice in all hematopoietic lineages. In another report, Lieber's group used helper-dependent adenoviral vectors to deliver PE5max prime editor in human HSPCs derived from SCD patients (Li *et al*, 2023). HSPCs i*n vitro* transduction resulted in 10% and 30% prime editing efficiency in absence or presence of transduced cell selection respectively, maintaining low level of indels (<1%).

3.5.2.2 Prime editing limitations and improvements

Despite prime editing represents a promising technology in the context of cutting-edge gene editing some limitations are still to be addressed: (i) limited efficiency, (ii) delivery, (iii) limited genome accessibility and (iv) byproducts and off target:

(i) The primary determinant of limited efficiency is the design of the pegRNA. Specifically, the spacer sequence, as well as the lengths of the PBS and RT template, strongly influence efficiency of editing. Efficient PBSs typically fall within a range of 8-15 nucleotides (nts), while optimal RT templates range from 10-20 nts (Anzalone *et al*, 2019a). However, these parameters depend on the sequence context and hardly predictable so far. Several software and artificial intelligence-based algorithms are available to design and help predicting the optimal pegRNA configurations for the target site and edit of interest (Chow *et al*, 2020; Hsu *et al*, 2021; Mathis *et al*, 2023). A significant breakthrough in the field was the development of engineered pegRNAs (epegRNAs), which improved PE efficiency by three-fold in cell lines. epegRNAs include an RNA motif at the 3' end to prevent degradation and enhance stability. An additional 8-nts linker connects the RNA motif to the PBS, in the attempt to prevent potential interference with pegRNA function. (Nelson *et al*, 2021).

(ii) Another challenge is represented by the size of PE, which is encoded by an approximately 6.5 kb gene. Despite the easiest way to deliver such an editor is through plasmid DNA, this approach proves to be toxic to HSPCs and is unsuitable for translational purposes (Wiehe *et al*, 2007). An alternative

approach is the use of viral vectors to deliver PE. However, the cargo capacity of most viral vector systems still limits their applicability. To overcome this limitation, dual-LV or dual-AAV delivery can be exploited, taking advantage of splicing proteins, such as inteins, to split the editor construct (Zhi *et al*, 2022; Anzalone *et al*, 2019a). To ensure transient expression of the editor, which would limit cell exposure and the risk of off-target activity, *in vitro* transcribed mRNAs can be used (Chen *et al*, 2021c). Similarly, RNP complexes could potentially facilitate short-term delivery, yet they have been tested in human cells with only modest efficiency (Petri *et al*, 2022b). Moreover, the absence of a commercially available PE protein limits the use of RNPs, especially when aiming to correct human primary cells, for which high-quality reagents are mandatory to minimize toxicity. A possible approach to facilitate PE delivery is to split the nCas9 and RT resulting in similar efficiency compared to the full PE (Grünewald *et al*, 2023; Liu *et al*, 2022).

(iii) Another significant limitation, as discussed above for Cas9 and BE technologies, is the restricted genome accessibility imposed by the NGG PAM requirement. The adoption of PAM-flexible Cas9 variants has been suggested (Kweon *et al*, 2021), despite their application in human primary cells is still pending.

(iv) Although prime editing does not require DNA DSB at the target sequence and bypasses HDR requirements, indels at the target site have been reported (Anzalone *et al*, 2019a). Indels are not only restricted to pegRNA target site but also found at the nicking gRNA site in PE3 or PE5 systems (Schene *et al*, 2020). Furthermore, in PE3 and PE5 systems the concomitant presence of pegRNA-mediated nick and nicking gRNA-mediated nick on the two opposite strands may result in loss of the genomic portion between them, thus pushing towards the requirement of further studies on the DSB byproducts at the target site. Regarding the pegRNA-dependent off-target site, they were found to be lower compared to gRNA-dependent off target in the context of nuclease Cas9 editing (Anzalone *et al*, 2019a). Indeed, the requirement of the hybridization steps of PBS and RT template in addition to the spacer-protospacer binding might reduce off-targets rate. More studies are required in order to evaluate

other type of off-target activities such as those induced by constitutive, albeit transient, expression of the RT.

4 AIMS OF THE WORK

Gene editing holds great promise for the treatment of inherited diseases by precise genetic engineering. The development of innovative editing systems, such as BEs and PEs, have expanded scope and means of gene editing (Rees & Liu, 2018a; Anzalone *et al*, 2019a). These cutting-edge gene editing systems may bypass some of the limitations related to nuclease-based gene editing in HSPCs and are considered a viable alternative to the latter for specific applications. Despite their promise for less harmful and more precise genetic engineering, little is known about short- and long-term toxicity of BEs and PEs in human cells. Concerns include conversion of the DNA SSB into a DSB during cell replication, expression of constitutive deaminases/RT that may have gRNA-independent genome-wide mutagenic potential, and adverse cellular responses triggered by the reagents and by processing nucleic acid intermediates of base and prime editing activity. Here, we perform a comparative assessment of state-of-the-art BEs and PEs versus Cas9 editing in HSPCs in terms of efficiency, cytotoxicity, detrimental cellular responses and on-target and genome-wide genotoxicity.

BEs and PEs are currently limited to short-range edits, particularly in primary cells. Targeted integration of large therapeutic sequences in HSPCs for gene correction or gene addition purposes requires efficient nuclease-based and HDR-based gene editing. Despite optimization of HDR editing protocols allows to reach 20 to 50% editing in long-term engrafting HSPCs (Lattanzi *et al*, 2021; Ferrari *et al*, 2020a) these efficiencies may fall short if edited cells do not bear selective advantage over the unedited counterpart and/or high proportion of functional cells is required to reverse the pathological process. Additionally, HDR-based techniques raise safety concerns due to the generation and processing of DNA DSB at the target site, leading to unintended and potentially genotoxic byproducts such as large deletions, translocations, chromothripsis, and integration of fragmented or full-length viral DNA (Adikusuma *et al*, 2018; Park *et al*, 2022; Turchiano *et al*, 2021; Leibowitz *et al*, 2021; Ferrari *et al*, 2022a). In parallel, we thus aimed to develop strategies coupling successful editing of a target gene with transient expression of a selector molecule to enrich for HSPCs bearing the intended HDR edit and purge out those carrying imprecise DNA DSB repair byproducts at the target site.

5 RESULTS

5.1 Base editing leads to imprecise outcomes at target sites ascribed to DNA DSBs

To side-by-side compare different editors, we selected a state-of-the-art version of CBE (BE4max) (Koblan *et al*, 2018b) and ABE (ABE8.20-m) (Gaudelli *et al*, 2020b) and used a gRNA targeting beta-2 microglobulin (*B2M)* (Gaudelli *et al*, 2020b) that can be coupled with either BE or Cas9 to induce its knock-out (KO). Since B2M is ubiquitously expressed on the cell surface, its KO allows straightforward quantification of editing efficiency by measuring lack of B2M expression via flow cytometry. Specifically, to induce *B2M* KO, Cas9 and BE4max introduce indels or a premature stop codon, respectively, while ABE8.20-m disrupts a splicing donor site (**Fig. 11a**). We took advantage of BE4max and ABE8.20-m commercially available plasmids, we added a post-transcriptionally regulatory element (WPRE) and a polyadenylation signal to increase stability and we produced BE mRNAs through *in vitro* transcription (**Fig. 11b**). In a B-lymphoblastoid cell line, which expresses B2M at high levels on the membrane, genetic modifications reached 40% with original plasmids, while mRNAs electroporation resulted in 60% B2M KO at the highest dose, as assessed by flow cytometry (**Fig. 11c,d**). As expected, *B2M* KO was not detected upon BE combination with an unrelated gRNA. Growth curve of cells revealed that toxicity was mostly ascribed to electroporation procedure and higher when delivering plasmid compared to mRNAs (**Fig. 11e**). Flow cytometry and molecular analysis of single cell-derived clones revealed that only biallelic KO reduced B2M expression on the cell surface, while monoallelic edited clones retained the same B2M expression level as wild type (WT) ones (**Fig. 11f**). Then, we purified mRNAs with high performance liquid chromatography to remove contaminants and reduce their toxicity in primary hematopoietic cells. In human primary T cells, BE4max and Cas9 resulted in 80-90% B2M KO cells, while ABE8.20-m reached >95% (**Fig. 11g**). All editing treatments showed comparable acute toxicity, mostly ascribed to electroporation (**Fig. 11h,i**). Overall, these data confirm the feasibility of BEs mRNA delivery.

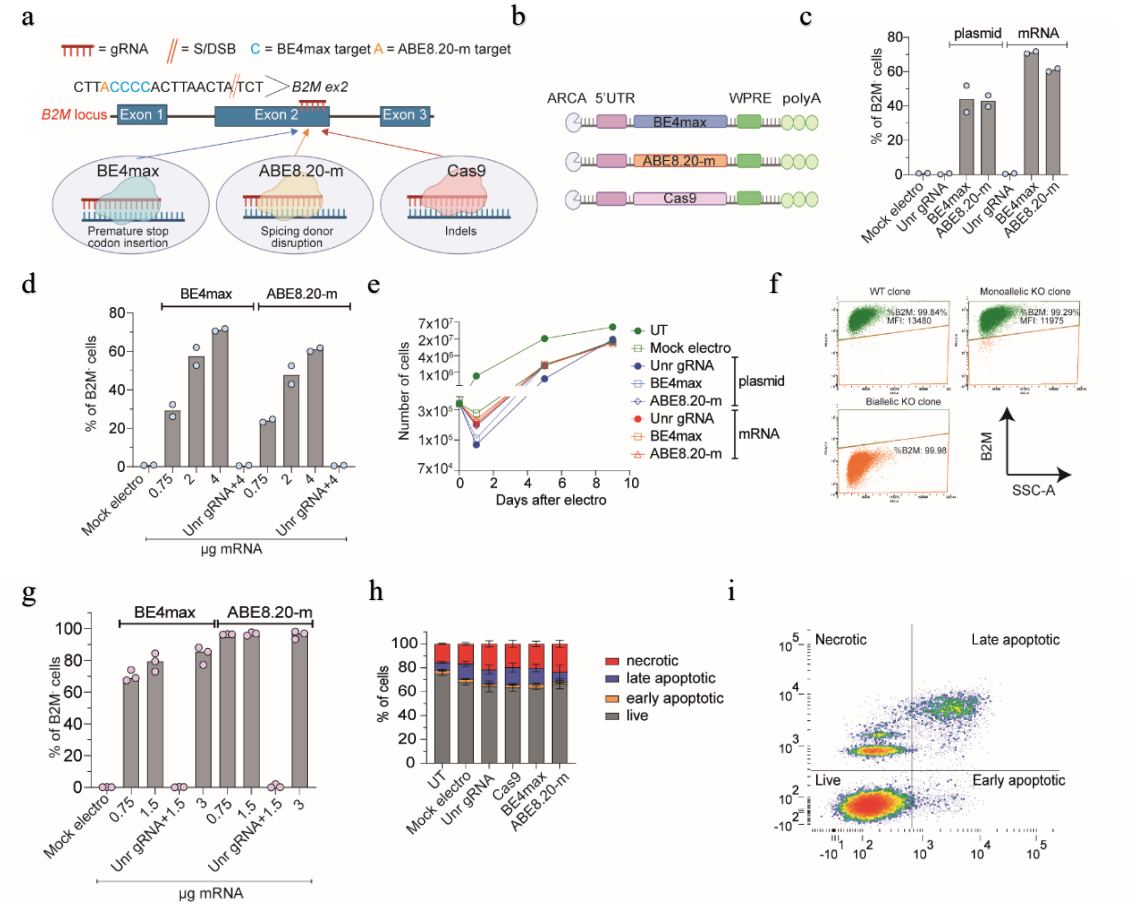


Figure 11. Base editing in cell lines and primary T cells. a,b, *Schematic representation of the B2M exon 2 editing strategies and their cognate genetic outcomes (**a**) and of the editor mRNAs (**b**). ARCA: Anti-Reverse Cap Analog; UTR: untranslated region; WPRE: Woodchuck Hepatitis Virus Post-transcriptional Regulatory Element.* ***c,d,*** *Percentage of B2M⁻ B-lymphoblastoid cells as measured by flow cytometry 7 days after editing (n=2). Unr: unrelated. Median.* ***e,*** *Growth curve of B-lymphoblastoid cells after treatments (n=2). Median.* ***f,*** *Flow cytometry plots of three representative B-lymphoblastoid clones showing wild-type (WT), monoallelic and biallelic editing confirmed by Sanger sequencing. SSC-A: Side Scatter.* ***g,*** *Percentage of B2M⁻ human T cells 7 days after treatments (n=3). Median.* ***h,*** *Percentage of live, early/late apoptotic and necrotic T cells 24 hours (hrs) after treatments. UT: untreated (n=3). Mean ± s.e.m.* ***i,*** *Representative plot showing gating strategy for live, early/late apoptotic and necrotic T cells.*

In cord blood (CB)- and mobilized peripheral blood (mPB)-derived CD34+ HSPCs we investigated different timings for gene editing, comparing 1 versus 3 days of culture after thawing (**Fig. 12a**). Whereas a longer protocol promotes metabolic activation and cell cycle progression, a shorter one may better preserve stem cell phenotypic markers. Indeed, since BEs do not require cell cycle activation and HDR engagement, we reasoned to forestall the editing protocol to preserve stem cell phenotype. We started by optimizing BE4max, ABE8.20-m and Cas9 mRNAs dose and we selected by flow cytometry and

molecular analysis the best performing one (7.5ug) for further studies. B2M KO was highly efficient for all systems, with ABE8.20-m outperforming BE4max and Cas9 reaching up to 88% and 64%, respectively, at the highest dose (**Fig. 12b-e**), without detectable changes in the proportion of different progenitor subsets (**Fig. 12e,f**). KO was lower at day 1 compared to day 3, in particular for BE4max, Cas9 and the most primitive progenitor subset. HSPCs treated with BE4max and ABE8.20-m showed similar *in vitro* clonogenic potential to mock electroporated cells and higher than Cas9-treated cells, pointing to a milder impact of BEs than Cas9 on HSPC function (**Fig. 12g**).

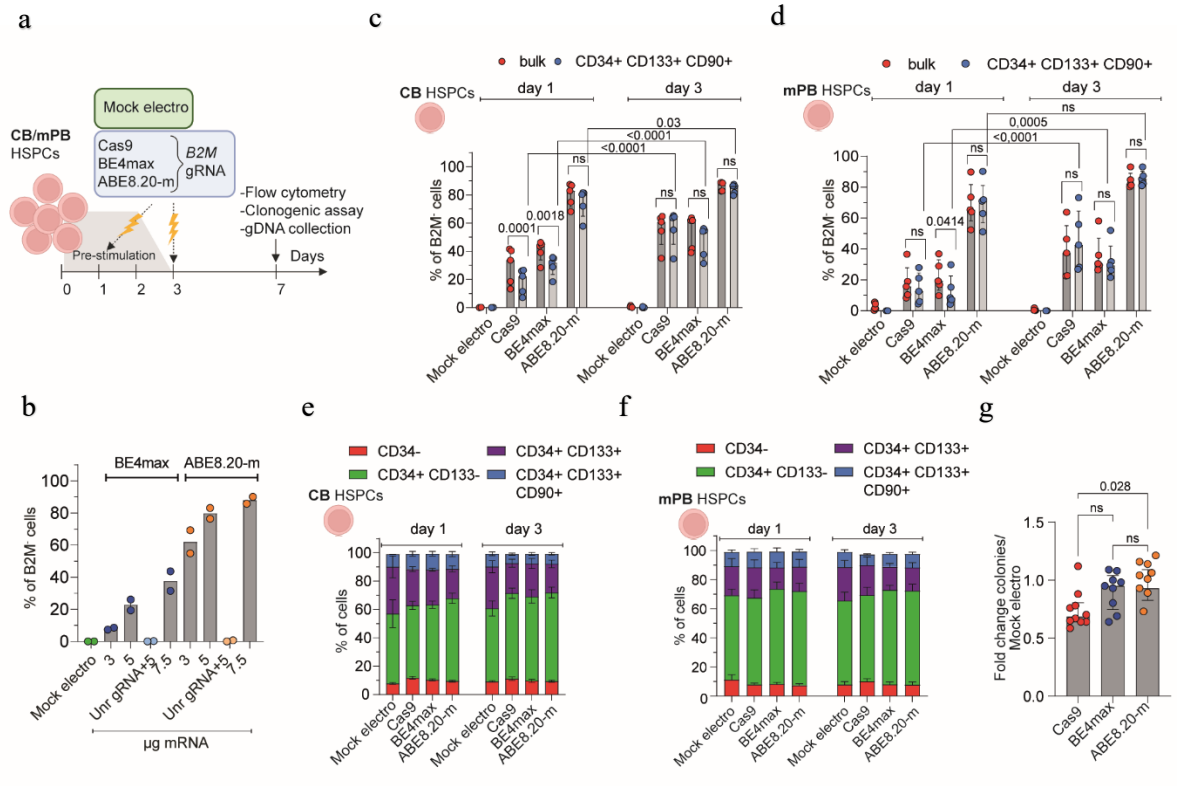


Figure 12. CB- and mPB-derived HSPCs base editing. a, *Experimental workflow for B2M editing in CB or mPB HSPCs.* ***b,*** *Percentage of B2M⁻ CB HSPCs after editing with different mRNA doses (n=2). Median.* ***c,d*** *Percentage of B2M⁻ CB (c) or mPB (d) HSPCs edited at day 1 or day 3 post-thawing (n=5). Median with interquartile range (IQR). Linear mixed-effects model (LME) followed by post hoc analysis.* ***e,f*** *Proportion of cellular subpopulations within CB (e) or mPB (f) HSPCs from experiments in 'c' and 'd' (n=5). Mean ± s.e.m. LME followed by post hoc analysis.* ***g,*** *Fold change in the number of colonies generated by CB or mPB HPSCs over mock electroporation (n = 10). Median with IQR. Kruskal-Wallis with Dunn's multiple comparison.*

We then sequenced the *B2M* target site from the edited cells of Fig. 12c,d and found the expected transitions at one or more target bases within the editing window in a proportion of alleles consistent with the fraction of biallelic KO reported above (**Fig. 13a,b**). However, while nearly all ABE8.20-m edited alleles showed base transitions, more than one third of BE4max-edited alleles carried indels at the target site. Whereas Cas9-induced indels spanned around the expected DNA DSB site, BE4max indels mostly occurred between the expected nCas9 and BE target sites (**Fig. 13c**). The fraction of indels-bearing alleles was higher for BE4max editing at day 1 than 3, when expression of several BER genes, such as *APEX1* and the upstream sensor *UNG*, was also higher (**Fig. 13d,e**). These findings suggested that excess indels induced by BE4max editing might be due to insufficient UG inhibition by the UGI domain (Komor *et al*, 2016) and the combined action of the BER-dependent APEX1 endonuclease and nCas9 to generate a DSB at the target sequence. Of note, some SNVs other than the expected transitions ("Other" in Fig. 13a,b) were also found at the target locus for both BEs, as also reported in other studies of CBE (Kurt *et al*, 2021a; Zeng *et al*, 2020c), suggesting occasional and/or supplemental engagement of alternative repair pathways.

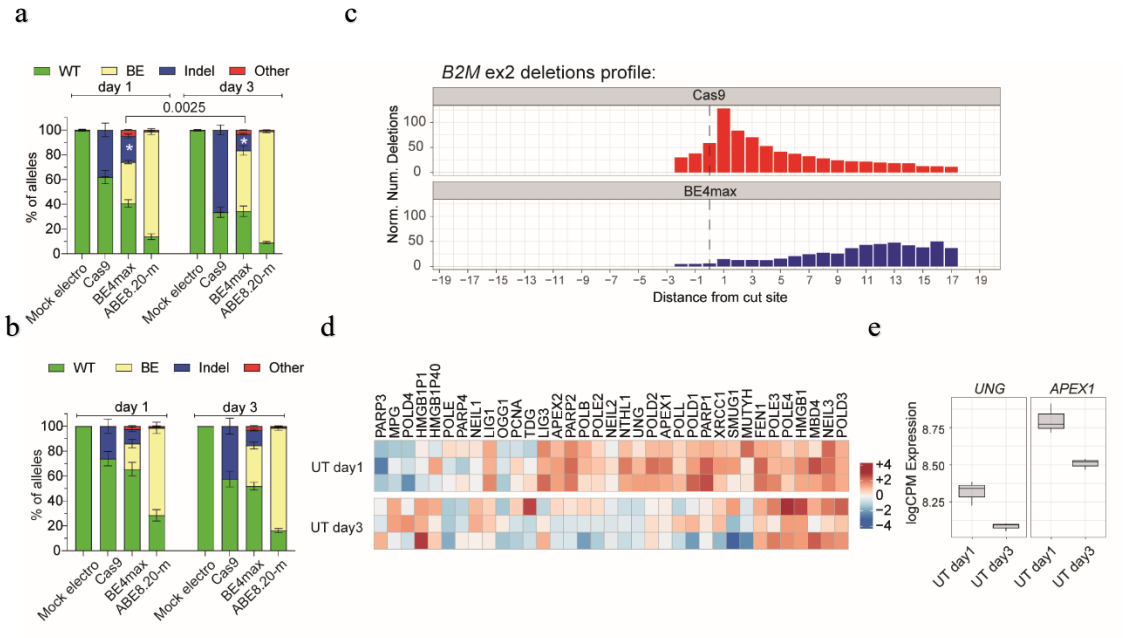


Figure 13. Dissecting editing outcomes at the target locus. a, *Percentage of B2M alleles, measured by deep sequencing, being WT or carrying the described editing outcomes in CB HSPCs (n=5 for day 1; n=6,7,7,7 for day 3). Mean ± s.e.m. Mann-Whitney test.* ***b,*** *Percentage of B2M alleles, measured by deep sequencing analysis, being WT or carrying the described editing outcomes in mPB HSPCs*

(n=4,5,5,5 for day 1; n=3,4,4,4 for day 3). Mean ± s.e.m. ***c,*** *Distribution of the distance of indels from the B2M exon 2 S/DSB cut site in Cas9-edited (top) and BE4max-edited samples (bottom).* ***d,*** *Heatmap of normalized read counts for genes belonging to the BER pathway (KEGG database; hsa03410) in UT CB HSPCs cultured for 1 or 3 days (n=3).* ***e,*** *UNG and APEX1 log counts per million reads (CPM) in UT CB HSPCs cultured for 1 or 3 days (n=3). Centre of the boxplot represents median, and boundaries represents first and third quartiles. Upper and lower whiskers extend 1.5xIQR from the hinge.*

To provide a broader representation of target sequence composition, including for the number and position of editable bases, additional gRNAs targeting the genomic safe harbor Adeno-Associated Virus Site 1 (*AAVS1*), exon 1 of *B2M* and the therapeutically relevant *BCL11A* erythroid-specific enhancer, *CCR5* and *IL2RG* were selected (**Fig. 14a**). At nearly all tested loci, ABE8.20-m outperformed BE4max in terms of efficiency and precision at the target site (**Fig. 14b**). Although indels and other unexpected SNVs were relatively frequent and more common for BE4max, some indels were also retrieved for ABE8.20-m, in particular when targeting exon 1 of *B2M*. The higher occurrence of indels at the latter site allowed describing the deletions profile and revealed a distribution centered on the gRNA cut site, similarly to Cas9, implying a different mechanism from that postulated above for BE4max, and likely due to conversion of SSB to DSB upon DNA replication (**Fig. 14c**). Consistently with the lower proportions of indels, the fraction of primitive HSPCs was not affected by ABE8.20-m treatment, while it was significantly decreased upon BE4max and Cas9 treatments (**Fig. 14d**).

Overall, these data show that highly efficient base editing may lead to imprecise outcomes at the target sites, comprising the genetic scars of repaired DSBs. These events are exacerbated in the case of BE4max by its interaction with the BER pathway.

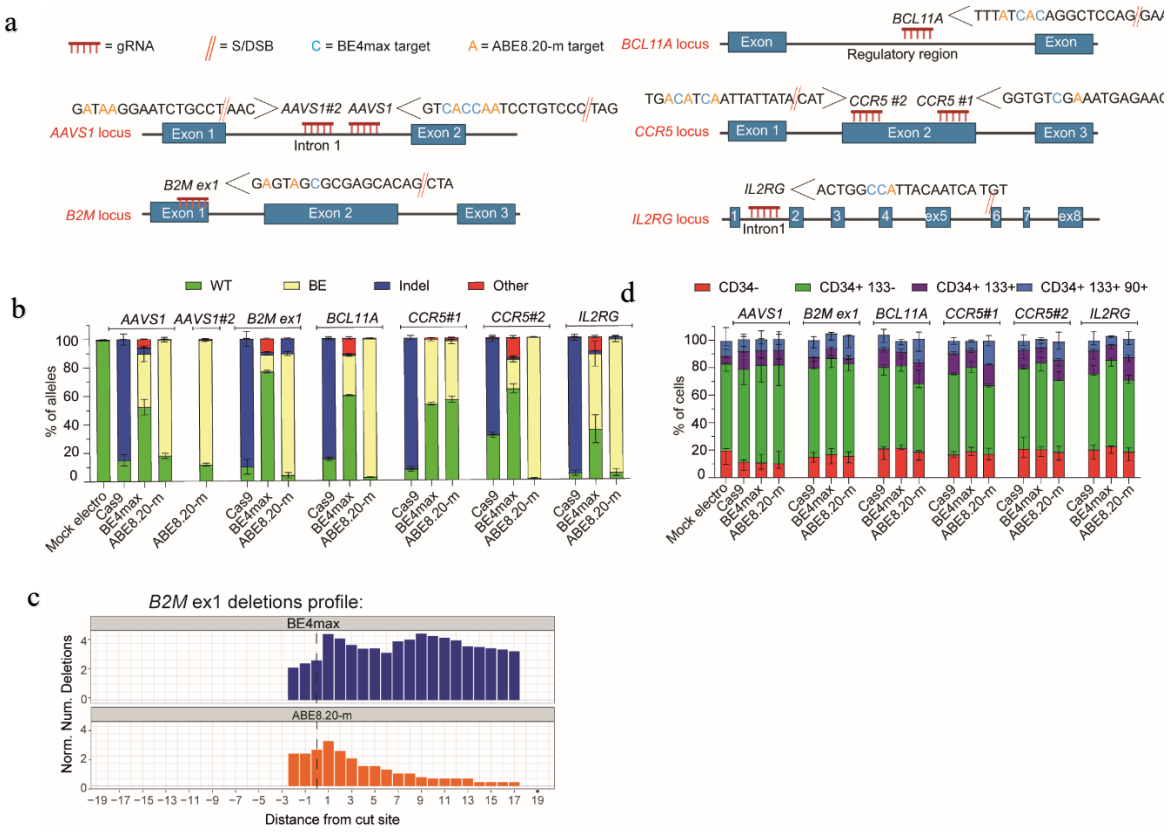


Figure 14. HSPCs base editing across different loci. ***a,*** *Schematic representation of the AAVS1, B2M exon 1, BCL11A, CCR5 and IL2RG editing strategies.* ***b,*** *Percentage of AAVS1, B2M exon 1, BCL11A, CCR5 and IL2RG alleles, measured by deep sequencing, being WT or carrying the described editing outcomes in mPB HSPCs (n=3 for AAVS1 Cas9; n= 7 for AAVS1 BE4max and ABE8.20-m; n=3 for B2M exon 1, BCL11A, CCR5 and IL2RG). Mean ± s.e.m.* ***c,*** *Distribution of the distance of indels from the B2M exon 1 S/DSB cut site in BE4max-edited (top) and ABE8.20-m-edited samples (bottom).* ***d,*** *Proportion of cellular subpopulations within mPB HSPCs from experiments in 'b' (n=3). Mean ± s.e.m. Samples edited in BCL11A, CCR5 and IL2RG were unified for statistical analysis using Friedman test with Dunn's multiple comparison on the most primitive compartments (CD34+ CD133+ and CD34+ CD133+ CD90+), as experiments were performed in parallel on the same mPB HSPC donors. Cas9 and BE4max showed a significant reduction in the proportion of primitive compartments compared to ABE8.20-m (0.016 and <0.0001 respectively).*

5.2 Base editing does not abrogate large deletions and translocations at target sites

To more comprehensively evaluate the spectrum of genetic outcomes at target site of different editing systems, we screened ~300 randomly picked colonies from the outgrowth of BE4max, ABE8.20-m and Cas9 treated mPB-derived HSPCs for the occurrence of large deletions extending upstream or downstream the *B2M* exon 2 or exon 1 gRNA target sites (**Fig. 15a**). For *B2M* exon 2 targeting, we found mono- or, less

frequently, bi-allelic loss of the interrogated locus in ~12% Cas9 and ~5% BE4max but only rare ABE8.20-m colonies (**Fig. 15b**). Of note, a higher proportion of deletions was found when probing downstream the BE4max cut site, in line with the skewed indels pattern in Fig. 13c. For *B2M* exon 1 targeting, we found ~15% Cas9 and ~3% ABE8.20-m but only rare BE4max colonies, where the ABE8.20-m data are consistent with a high proportion of indels at this site and the BE4max data reflect a low editing efficiency (**Fig. 15c** and see Fig. 14b). We then probed for the possible occurrence of translocations between multiplex editing sites on different chromosomes by co-delivering two gRNA targeting *AAVS1* and *B2M* exon 2 together with each editing systems and amplifying interchromosomal junctions by a matrix of PCR primers binding to each side of both editing loci (**Fig. 15d**). As expected from the high rate of indels and large deletions, Cas9-treated samples were positive for all 4 possible translocation events between the two sites (**Fig. 15e-h**). Notably, translocations were also clearly detectable for BE4max samples in 2 out of the 3 tested donor for the interchromosomal junction #1 (**Fig. 15e**) and in 4 out of 6 tested donors for the interchromosomal junction #2 (**Fig. 15f**), but not for ABE8.20-m. Sanger sequencing of *B2M-AAVS1* junctions revealed that while Cas9 translocations originated precisely from the respective cut sites, BE4max translocations were more heterogenous and spanning from the predicted nCas or APEX1 nicking sites on either side of the junction (**Fig. 15i**).

Overall, these results highlight the occurrence of potentially genotoxic outcomes at BE target sites consequent to DNA DSBs, such as large (≥ hundreds bps) deletions and translocations, at rates lower than observed for Cas9 and consistent with the fraction of indels detected by targeted deep sequencing.

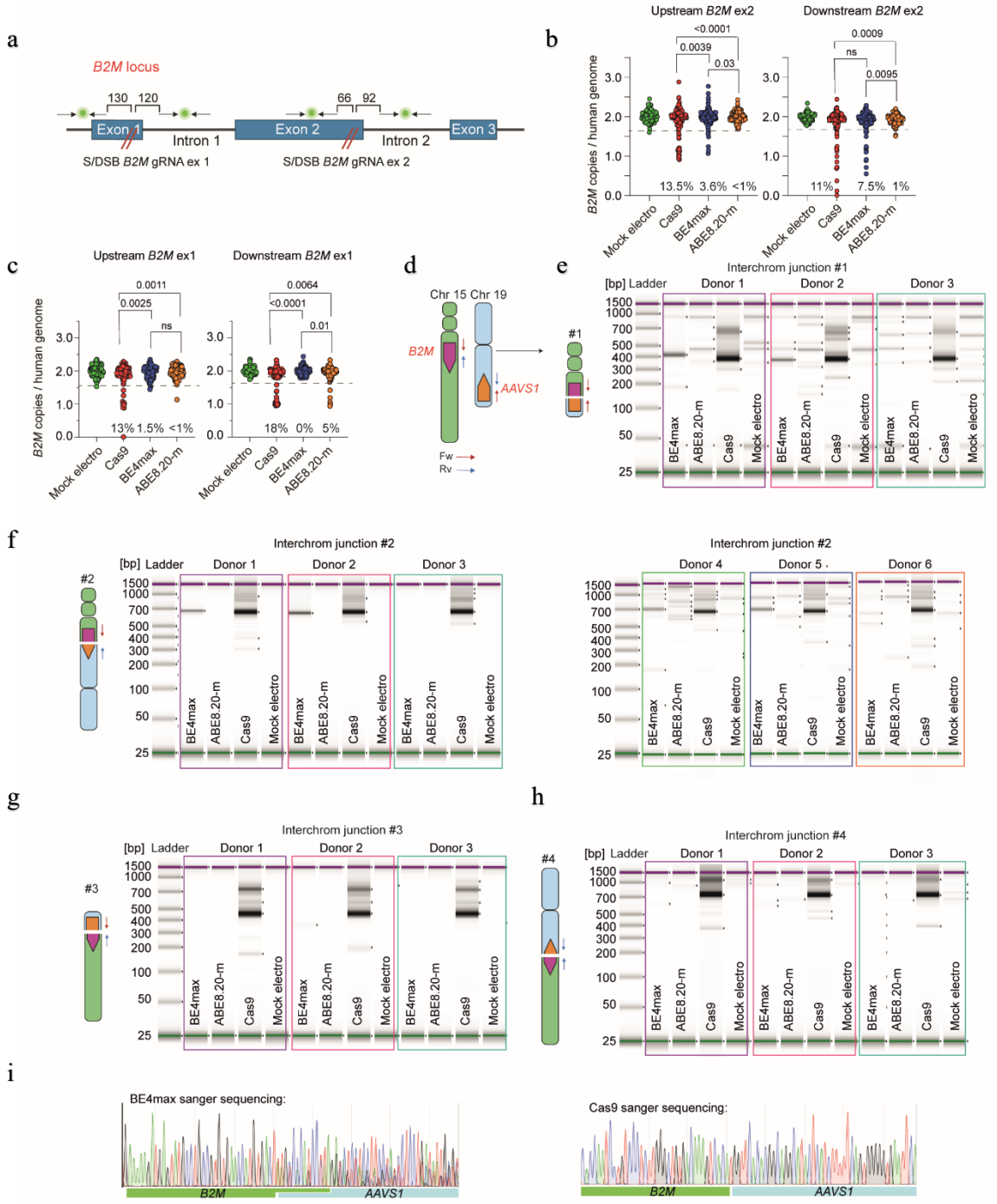


Figure 15. Assessing large deletions and translocation upon base editing. a, *Schematic representation of the probes used for deletions detection at B2M target sites in exon 2 and exon 1. The distances between the target site and the closest primer of the ddPCR amplicons are shown.* ***b,*** *Copies of B2M sequences per human genome flanking the exon 2 target site in individual colonies generated by edited mPB HSPCs (n=105,188,188,187 for 'upstream' assay; n=93, 188,188,187 for 'downstream' assay). Dashed lines indicate the lower limit of the confidence interval from 'Mock electro' colonies. Median. Fisher's exact test.* ***c,*** *Copies of B2M sequences per human genome flanking the exon 1 target site in individual colonies generated by edited mPB HSPCs (n=89,129,130,125 for 'upstream' assay;*

n=89,129,129,126 for 'downstream' assay). Dashed lines indicate the lower limit of the confidence interval from 'Mock electro' colonies. Median. Fisher's exact test. ***d,*** *Schematic representation of translocations expected upon multiplexed B2M and AAVS1 targeting.* ***e-h*** *Images of capillary electropherogram showing amplification of interchromosomal junction #1, #2, #3, #4 upon HSPC editing with two gRNAs targeting B2M exon 2 and AAVS1 in 3 (e,g,h) or 6 (f) mPB donors.* ***i,*** *Representative Sanger sequencing plot of B2M exon 2-AAVS1 junction in samples from 'f' edited with BE4max (left) or Cas9 (right).*

5.3 BEs trigger p53 activation and IFN response in HSPCs

We then investigated the cellular transcriptome 24 hrs after treatment with the different editors to identify detrimental responses that may impact HSPC function (**Fig. 16a**). Beside positive enrichment for genes belonging to apoptosis and inflammation categories in all samples due to electroporation *per se*, BE4max and Cas9 triggered p53 pathway activation (**Fig. 16b,c**), with upregulation of nearly identical set of genes, pointing towards sensing and repair of DNA DSB as the common trigger (**Fig. 16d**). The p53 response was lower for BE4max than Cas9, consistently with the above findings for indels and large deletions at the editing site but still raising concern for a detrimental impact on HSPC function. In addition, BE4max and ABE8.20-m, but not Cas9, activated interferon (IFN)-α and IFN-γ responses (**Fig. 16c**). Unbiased clustering of IFN-α and IFN-γ target genes revealed upregulated subsets upon BE treatments enriched for RNA recognition ontologies, possibly indicating innate cellular sensing of long mRNAs (~6kb; **Fig. 16e-g**).

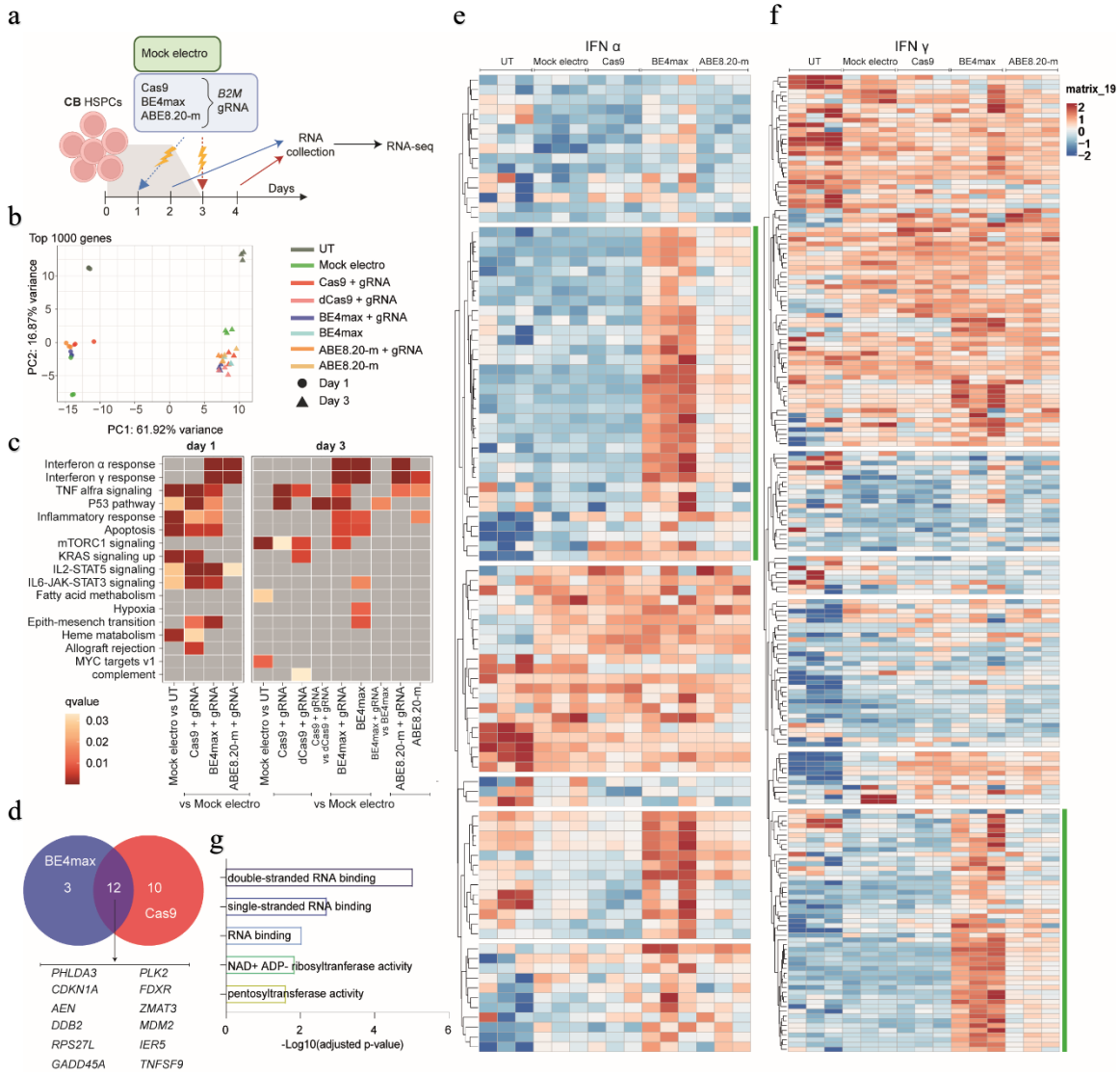


Figure 16. Evaluating cellular responses upon base or Cas9 editing. ***a,*** *Experimental workflow for B2M exon 2 editing in a pool of 6 CB HSPC donors for transcriptomic analysis (n=3 technical replicates for each condition).* ***b,*** *Principal component (PC) analysis from the RNA-Seq dataset in 'a'.* ***c,*** *Heatmap of q-values of enriched categories for selected comparisons between treatments on upregulated genes (FDR < 0.05 and logFC > 0). dCas9: catalytically inactive (dead) Cas9. Enrichment test.* ***d,*** *Venn diagram representing the number of p53 target genes upregulated after BE4max or Cas9 treatments.* ***e,f,*** *Heatmaps of normalized read counts for target genes belonging to IFN-α (e) and IFN-γ (f) response categories across samples. Green lines indicate the subset of genes identified by unsupervised clustering with higher normalized read counts upon BE treatments.* ***g,*** *Adjusted p-values for the top-5 enriched categories (Hallmark gene set) when computing genes belonging to the green cluster from 'e,f'. Enrichment test.*

5.4 BE4max but not ABE8.20-m impairs long-term engraftment of edited HSPCs

To investigate editing of the small fraction of repopulating HSPCs comprised within CB-derived CD34+ cells, we transplanted immunodeficient mice with the outgrowth of matched numbers of cells seeded in culture (day 0) and treated for *B2M* exon 2 editing by the different systems at day 1 or day 3 (**Fig. 17a,b**). Longitudinal PB sampling and bone marrow (BM) and spleen (SPL) analyses at the end of the experiments showed long-term engraftment and multilineage reconstitution by human cells in xeno-transplanted mice for all edited samples (**Fig. 17c,e,g,h**). Despite the longer culture time HSPCs edited at day3 engrafted long-term similarly to cells edited at day1, suggesting a lower impact of the editing procedure in cycling cells (**Fig. 17c**). Moreover, day3 edited cell displayed a faster hematopoietic reconstitution most likely due to engraftment of progenitor cells. Remarkably, ABE8.20-m and Cas9-edited cells displayed lower long-term engraftment capacity than mock electroporated controls but maintained stable editing efficiencies in PB and hematopoietic lineages, comparable to the *in vitro* ones (**Fig. 17d,f**). While BE4max edited cells showed comparable engraftment capacity to mock electroporated cells, the editing efficiency was lower when compared to the *in vitro* results and further decreased over time in the graft (from around 50% to 10%), suggesting an impact of the editor and/or a lower editing efficiency in the long-term engrafting fraction. Deep sequencing analysis of the *B2M* exon 2 target site in the human cells retrieved from the mice showed nearly exhaustive occurrence of the expected transitions for the ABE8.20-m samples and a lower proportion for the BE4max samples, consistent with the fraction of engrafted B2M KO cells (**Fig. 17i**). Indels accounted for most of editing in Cas9 engrafted cells, but also contributed considerably in the BE4max samples, where they were more abundant in cells edited at day 1 and decreased from early to late timepoints. The latter observation might correlate with higher BER genes expression in day-1 cultured cells and in multipotent or lineage-committed progenitors vs. primitive HSCs, as reported by some of us in a ss RNA-Seq analysis of CD34+CD133+ CB cells cultured for 4 days (Schiroli *et al*, 2019b) (**Fig. 17j,k**). Indels were much fewer but still present in ABE8.20-m samples, averaging 1-2%. Similar findings in terms of engraftment and editing efficiency were obtained when transplanting mPB-derived HSPCs edited at day 3 at *B2M* exon 2 (**Fig. 17l,m**).

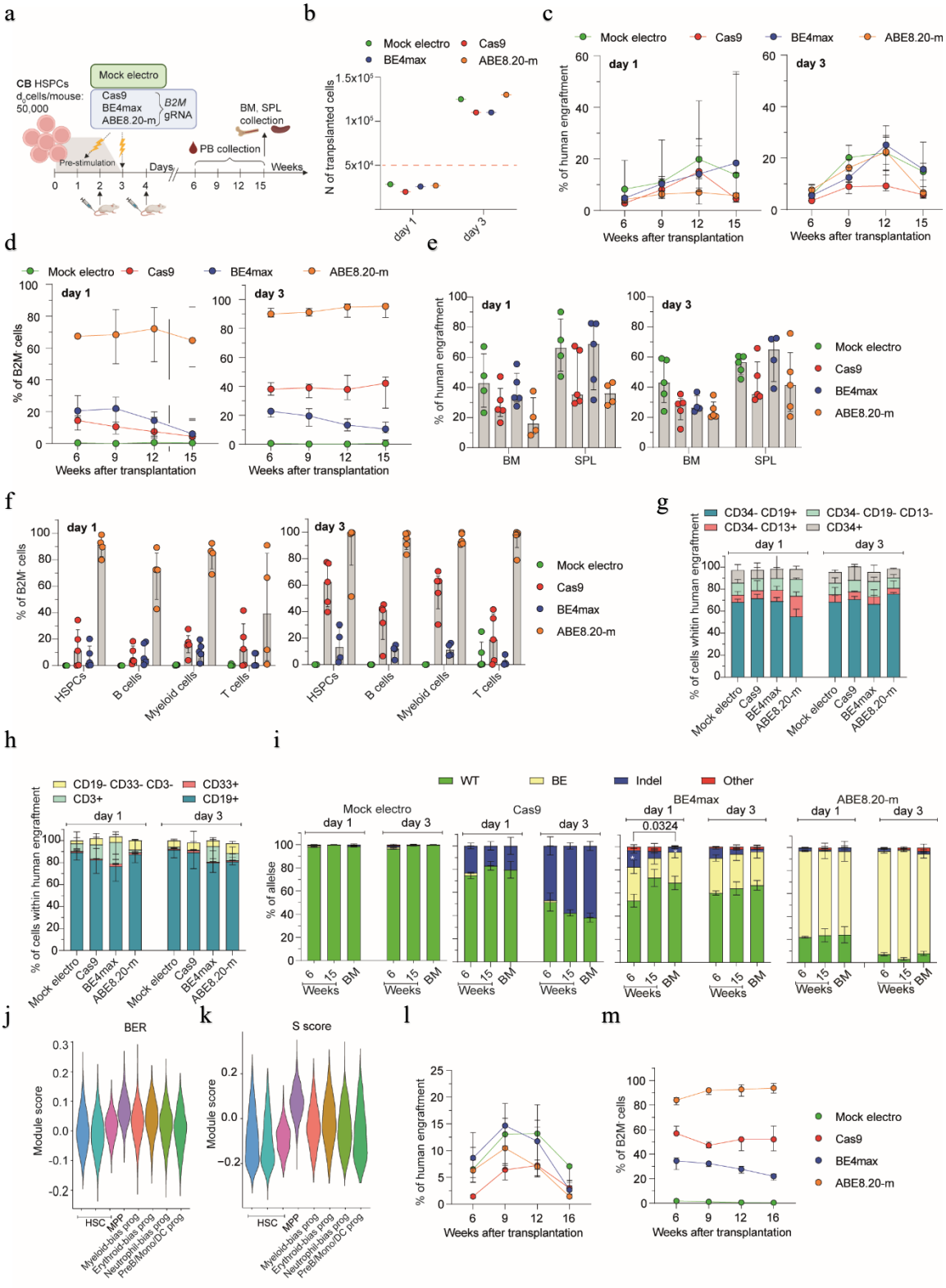


Figure 17. Evaluating in vivo persistence of B2M edited HSPCs. *a, Experimental workflow for B2M exon 2 editing in CB HSPCs and xenotransplantation. BM: bone marrow. SPL: spleen. b, Number of cells transplanted per mouse for experiment in 'a'. The red line corresponds to the* d_0 *equivalent.* ***c,d,*** *Percentage of human cells engraftment (c) and* $B2M^-$ *cells within*

human graft (d) in mice from 'a' transplanted at day 1 (left; n=5,5,5,4) or day 3 (right; n=5,5,4,5) post-thawing. Median with IQR. LME followed by post hoc analysis. ***e,f,*** *Percentage of human cells engraftment (e) and B2M⁻ cells (f)) in BM and SPL in mice from 'a' (n=5,5,5,4 for day 1; n=5,5,4,5 for day 3). Median with IQR.* ***g,h*** *BM (g) and SPL (h) lineage composition in mice from 'a' (n=5,5,5,4 for day 1; n=5,5,4,5 for day 3). Mean ± s.e.m.* ***i****, Percentage of B2M exon 2 alleles, measured by deep sequencing, being WT or carrying the described editing outcomes in mice from 'a' (Mock electro n=4 for day 1, n=5 for day 3; Cas9 n=5 day 1, n=5 for day 3; BE4max n=5 for day 1, n=4 for day 3; ABE8.20-m n=4 day 1, n=5 for day 3). Mean ± s.e.m. Kruskal-Wallis with Dunn's multiple comparison.* ***j,k*** *Module score for genes belonging to the BER pathway (KEGG database; hsa03410) (j) or S-phase signature from (Nestorowa* et al*, 2016) (k) in different HSPC subsets from (Schiroli* et al*, 2019a). MPP: multipotent progenitors.* ***l,m*** *Percentage of human cells engraftment (l) and B2M⁻ cells within human graft (m) in mice transplanted with mPB HSPCs edited at B2M exon 2 at day 3 post-thawing.*

To confirm our findings, we then transplanted mPB-derived HSPCs edited in *AAVS1* locus at day 3 (**Fig. 18a,b**). PB analyses revealed a similar pattern compared to the previous experiment in term of human engraftment at early timepoints, with Cas9 edited cells showing lower engraftment potential compared to the other treatments (**Fig. 18c**). Editing overtime was stable for Cas9 and ABE8.20-m treatment, with the latter reaching ~100% modified alleles (**Fig. 18d**). In contrast, BE4max editing progressively declined thus confirming the trend of previous experiments. Endpoints analyses on hematopoietic organs confirmed a lower human engraftment of Cas9 edited cells compared to BE and mock electro treatments (**Fig. 18e**), and the lowest editing efficiency in BE4max treated mice (**Fig. 18f**).

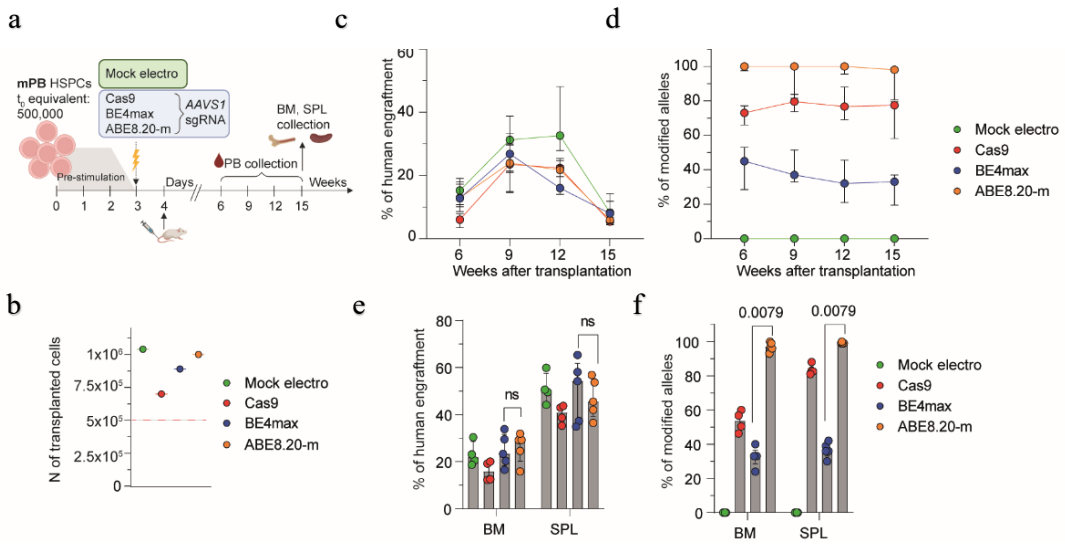


Figure 18. Evaluating in vivo persistence of AAVS1 edited HSPCs. a, *Experimental workflow for AAVS1 editing in mPB HSPCs and xenotransplantation.* ***b,*** *Number of cells transplanted per mouse for experiment in 'a'. The red line corresponds to the d_0 equivalent.* ***c,d*** *Percentage of human cells engraftment (c) and AAVS1 modified alleles within human graft (d) in mice transplanted with mPB HSPCs after AAVS1 editing at day 3 post-thawing (n=4,4,5,5). Median with IQR. LME followed by post hoc analysis.* ***e,f,*** *Percentage of human cells engraftment (e) and AAVS1 modified alleles (f) in BM and SPL from 'a' (n=4,4,5,5). Median with IQR. Mann-Whitney test.*

To investigate more stringently whether base editing could affect the output of single HSPC clones, we tracked cells treated by BEs or Cas9 at the clonal level. Because it is hardly possible to couple BE to a unique genetic identifier, we transduced HSPCs with a LV carrying a reporter (truncated low-affinity Nerve Growth Factor Receptor, dNGFR) and a degenerated barcode sequence (BAR) prior to editing and xenotransplantation (**Fig. 19a,b**). HSPCs transduction after 1 day of pre-stimulation resulted in vector copy number of 1.5 and 60% reporter expression in the bulk population, thus confirming marking of most HSPCs (**Fig. 19c**). However, since marked HSPCs were then edited at *B2M* locus at day 3 this strategy cannot discriminate between edited and non-edited cells in the graft. In accordance with previous data, the graft size was reduced for Cas9 samples (**Fig. 19d**), while editing efficiency reached ~40% for Cas9 and BE4max and ~100% for ABE8.20-m (**Fig. 19e**). Hence, an altered behavior of ABE8.20-m edited cells should be easily captured by interrogating the whole graft, while this may not apply to Cas9 and BE4max treated cells. While Cas9 treatment led to a moderate shrinkage of clonal complexity in the hematopoietic organs and in most of the sorted hematopoietic lineages, consistently with the lower graft size, ABE8.20-m and BE4max did not show reduced clonality compared to mock electroporated control (**Fig 17f**). Since 100-150 repopulating clones were retrieved from base edited and mock samples, we calculated a frequency of 1 out of 1.0-1.5x10^3 transplanted cells (**Fig. 19g**), which is in the range of previously reported findings from limiting dilutions transplant of uncultured CB-derived HSPCs (Bai *et al*, 2019b) validating our analysis and highlighting no significant loss of long-term engrafting clones upon BE treatments. Longitudinal analysis of PB revealed the progressive disappearance of some short-term engrafting clones and emergence of long-term engrafting ones independently of the treatment, in line with previous observations on HDR-edited xenografts (Ferrari *et al*, 2020b) and gene therapy patients (Scala *et al*, 2018) (**Fig. 19h**).

When computing the sharing index among samples retrieved from each individual mouse, we found the highest scores among B, myeloid and HSPCs lineages and the lowest for T cells, possibly indicating an independent origin for lymphoid-biased progenitor (**Fig. 19i**). Importantly, 75% of BARs retrieved from CD34+ HSPCs derived from BE4max, ABE8.20-m, Cas9 or control mice were shared with ≥2 differentiated hematopoietic lineages, thus confirming the long-term multilineage repopulation capacity of individual HSPCs treated for editing (**Fig. 19j**). Whereas most BARs were predominantly shared across the different lineages of the same mouse, some were also shared among different mice, more often within the same treatment group (**Fig. 19k**).

Overall, these findings show that ABE8.20-m efficiently edits and preserves multi-lineage output of long-term repopulating HSPCs, while BE4max is less efficient and adversely impact repopulation by edited cells.

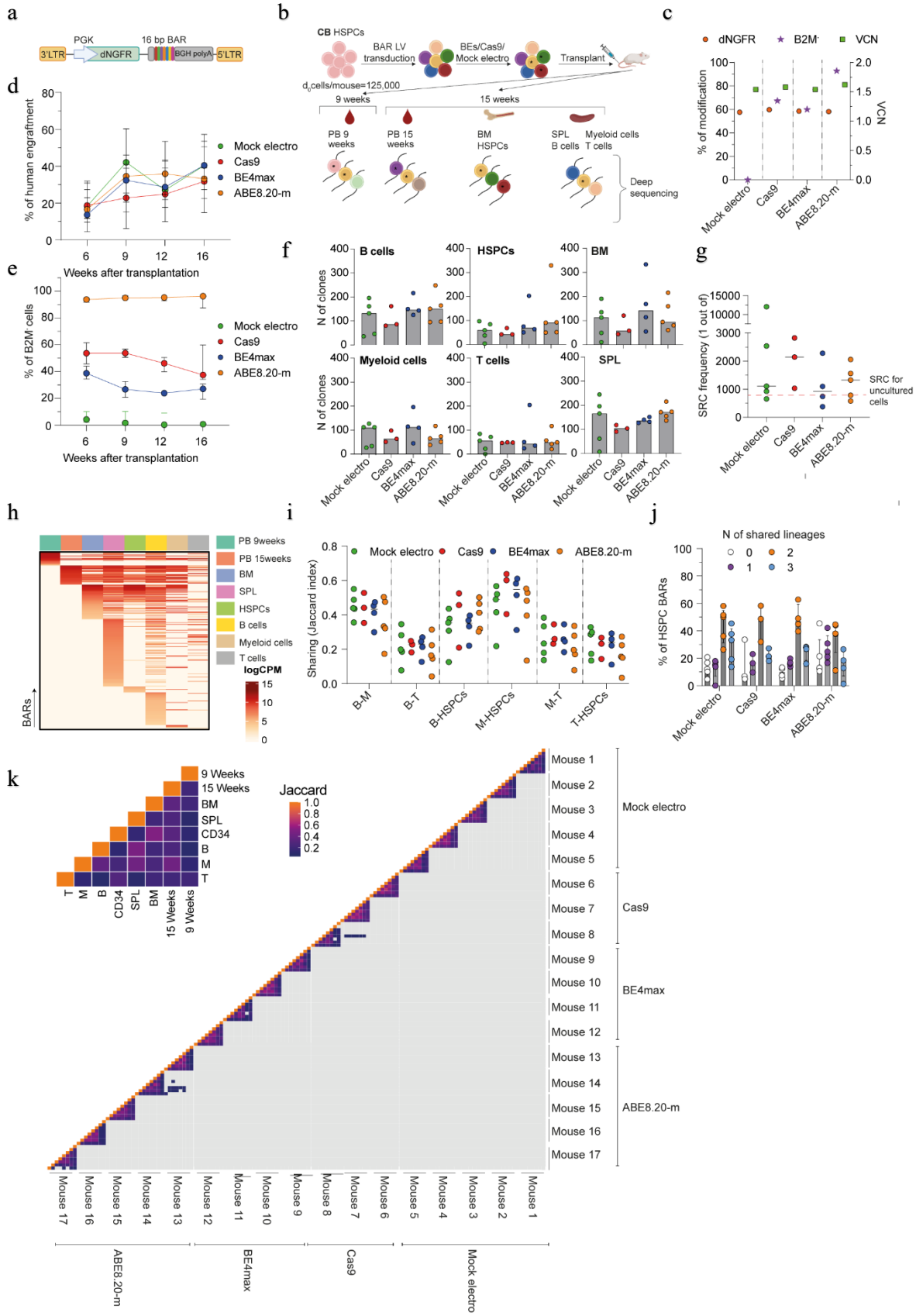

a
PGK
16 bp BAR
3'LTR
dNGFR
BGH polyA
5'LTR
b
CB HSPCs
BAR LV transduction
BEs/Cas9/ Mock electro
Transplant
d0cells/mouse=125,000
9 weeks
15 weeks
PB 9 weeks
PB 15 weeks
BM HSPCs
SPL B cells
Myeloid cells T cells
Deep sequencing
c
dNGFR
B2M
VCN
% of modification
VCN
Mock electro
Cas9
BE4max
ABE8.20-m
d
% of human engraftment
Weeks after transplantation
e
% of B2M cells
Weeks after transplantation
f
B cells
HSPCs
BM
Myeloid cells
T cells
SPL
N of clones
g
SRC frequency (1 out of)
SRC for uncultured cells
h
PB 9weeks
PB 15weeks
BM
SPL
HSPCs
B cells
Myeloid cells
T cells
logCPM
BARs
i
Sharing (Jaccard index)
B-M
B-T
B-HSPCs
M-HSPCs
M-T
T-HSPCs
j
N of shared lineages
% of HSPC BARs
k
9 Weeks
15 Weeks
BM
SPL
CD34
B
M
T
Jaccard
Mouse 1
Mouse 2
Mouse 3
Mouse 4
Mouse 5
Mouse 6
Mouse 7
Mouse 8
Mouse 9
Mouse 10
Mouse 11
Mouse 12
Mouse 13
Mouse 14
Mouse 15
Mouse 16
Mouse 17
Mock electro
Cas9
BE4max
ABE8.20-m

Figure 19. Clonal tracking of B2M edited HSPCs. a,b, *Schematic representation of the barcoded LV library (a) and the workflow (b) for CB HSPCs clonal tracking experiment. LTR: long-terminal repeats. PGK: phosphoglycerate kinase promoter. BGH: bovine growth hormone polyadenylation signal.* ***c,*** *Percentage of dNGFR$^+$ and B2M$^-$ cells, measured by flow cytometry, and vector copy number (VCN), measured by ddPCR.* ***d,e,*** *Percentage of human cells engraftment (d) and B2M$^-$ cells within human graft (e) in mice from 'b' (n=5,3,4,5). Median with IQR.* ***f,*** *Number of clones in hematopoietic lineages and organs from mice in 'b' (n=5,3,4,5). Median.* ***g,*** *Severe combined immunodeficient (SCID)-Repopulating Cells (SRC) frequency in mice from 'b', calculated dividing the d0 equivalent cell number by the number of engrafted clones from BM in 'f'. The red line shows the SRC for uncultured HSPCs (Bai* et al*, 2019c) (n=5,3,4,5). Median.* ***h,*** *Heatmap of the abundance (red-scaled palette) of BARs (rows) for one representative BE4max mouse in PB at indicated times after transplant, hematopoietic organs and lineages (columns).* ***i,*** *Jaccard index as a measure of BAR sharing between hematopoietic lineages (n=5,3,4,5). Median.* ***j,*** *Percentage of unique HSPC BARs shared with none, 1, 2 or 3 hematopoietic lineages (n=5,3,4,5). Median with IQR.* ***k,*** *Heatmap showing the Jaccard index as a measure of inter- and intra-sample BAR sharing in mice from 'b'.*

5.5 Transcriptome and exome analyses uncover global effects of BEs on the mutational landscape

In base editing approaches, the overexpression of deaminases domain, albeit transient, might result in gRNA-independent off-target activity raising concerns about the biological consequences of these treatments. To address this issue, we next evaluated the mutational burden induced by BEs at both transcriptomic and genomic level. We measured the mutational burden in the transcriptome by variant calling analysis of the reads from RNA-seq experiment in Fig 16. We found a consistent albeit moderate increase of mutational load on the transcriptome of HSPCs edited at day 1 or day 3 by ABE8.20-m, as compared to all other treatments despite similar levels of editor expression (**Fig. 20a,c**). Moreover, we detected an increased fraction of all SNV types and not only of the expected A>G transition (**Fig. 20b**).

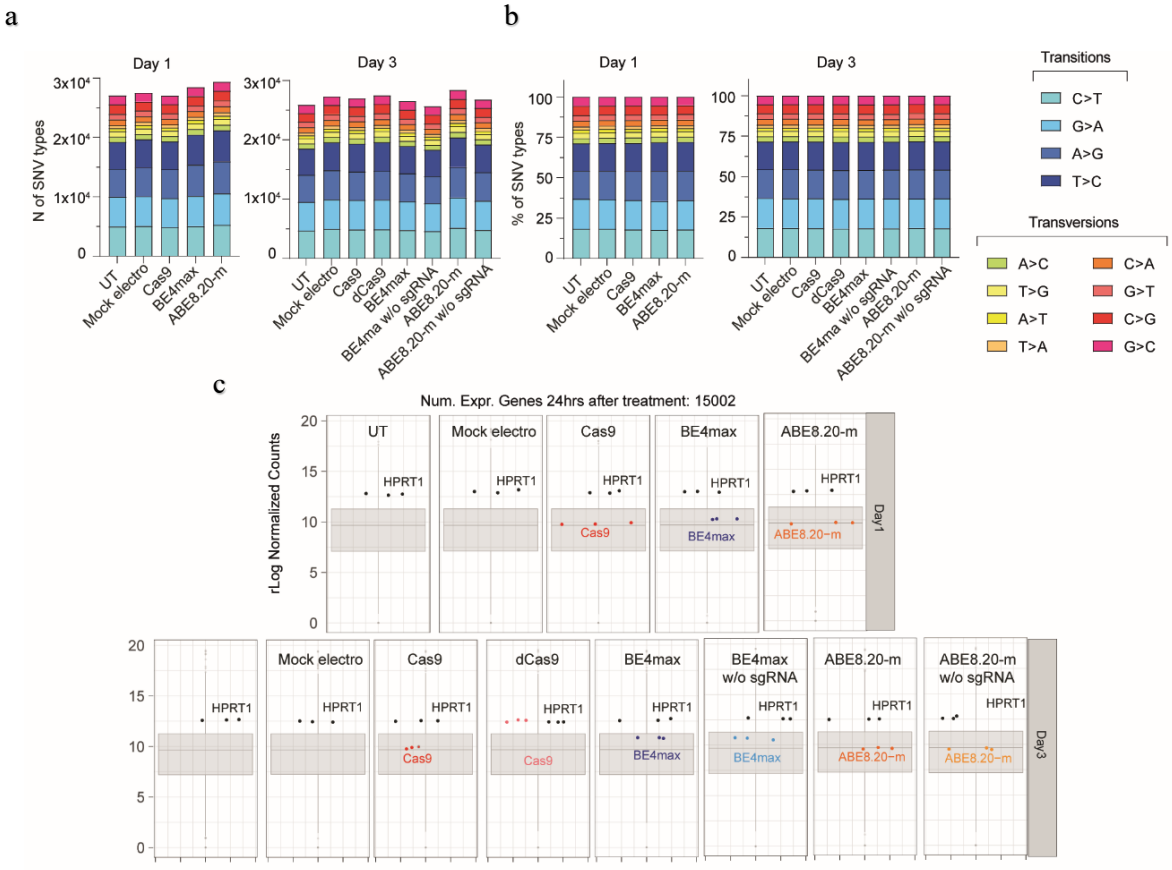


Figure 20. Mutational burden in transcriptome of B2M edited HSPCs. a,b, *Number of SNV types (a) and their relative proportions (b) in RNA-seq experiment in Fig. 16a.* ***c,*** *Boxplot showing the normalized expression (read counts) of the different editors and of the HPRT1 housekeeping gene in RNA-seq experiment in Fig. 16a.*

We then explored the possible occurrence of gRNA-independent genome-wide DNA mutagenesis on chromatin R-loops. We performed ultra-high coverage (500X) whole exome sequencing (WES) of the *in vitro* outgrowth of HSPCs treated with the different editors from the experiment described in Fig. 19**,** calling all variants against the reference human genome (GRCh38) and plotting their intersections among all samples (**Fig. 21a,b**). As expected, the vast majority of variants were shared by all samples, reflecting germline variants of the HSPC donors. We then subtracted all variants shared between mock electro

and ≥1 samples from each condition to capture treatment-associated variants and found that BE4max treatment increased their total amounts, but not the relative proportions of different SNVs, as compared to Cas9 or ABE8.20-m (**Fig. 21c-e**). We then postulated that analyzing the expanded clonal outgrowths contributing to the oligoclonal human hematopoietic graft of transplanted mice might increase the sensitivity of analysis towards detection of non-recurring genome-wide variants acquired by individual long-term repopulating cells during the treatment, albeit at the cost of limited sampling (**Fig. 21a**). The expected C>T or A>G transitions were highly represented at the *B2M* target locus in all BE4max and ABE8.20-m samples, respectively, thus validating our pipeline (**Fig. 21f**). Similarly, for Cas9 samples, we retrieved substantial proportions of indels in the target region, reflected by drops in reads alignment. For the genome-wide analysis, as before we subtracted all variants previously called for the mock electro sample in the *in vitro* analysis (see Fig. 21b) and computed those specific for each mouse/HSPC treatment. Remarkably, we found the lowest figures for BE4max samples, followed by ABE8.20-m, Cas9 and mock electro, in order of increasing numbers (**Fig. 21g**). This pattern was reminiscent of the impact of treatment on edited cell engraftment, as shown in Fig. 17d,m; Fig. 18d; Fig. 19e , with variant diversity being a proxy for clonality. However, when we computed the different types of variants and relative proportions of SNV types, we found a similar pattern among ABE8.20-m, Cas9 and mock samples, and a slight increase of indels and lower proportions of C>T/G>A transitions and higher proportions of A>C/T>G and G>C/C>G transversions in the BE4max samples, implying a treatment specific effect on the mutational landscape (**Fig. 21h,i**). When annotating high and moderate impact variants within a selected panel of cancer associated genes, which may provide selective advantage to mutant clones, we found few variants in all treated samples, most of which were shared among all treatment groups (**Fig. 21fj**).

In summary, at genome-wide level, treatment with BE4max showed alteration of the exome mutational landscape as compared to mock or other editing treatments, with an increased load in bulk-analyzed *in vitro* outgrowth of treated HSPCs and a substantial drop in the oligoclonal resulting graft. Notably, the latter observation was accompanied by a skewed distribution disfavoring the expected deaminase-induced transitions.

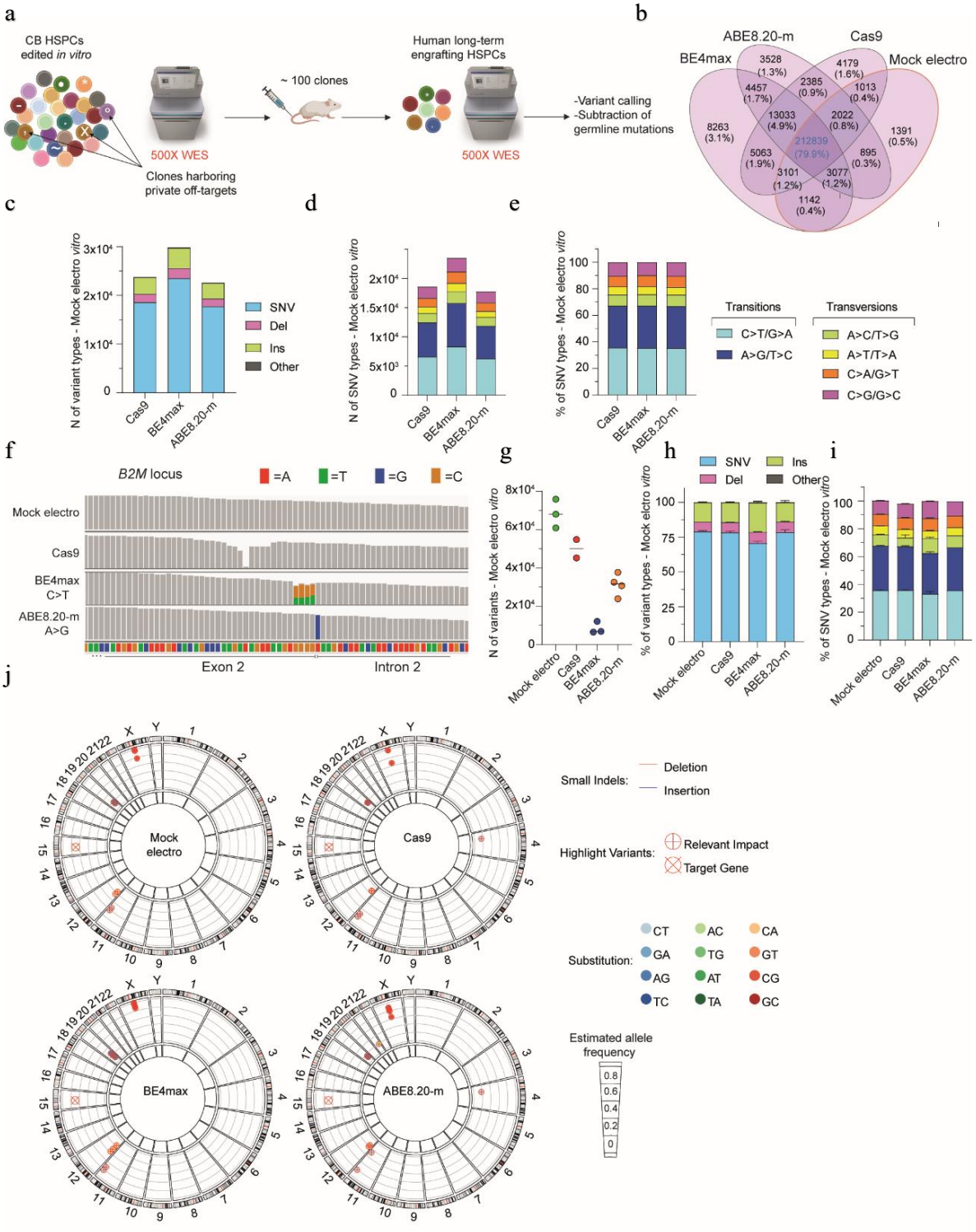


Figure 21. Mutational burden in exome of B2M edited HSPCs. a, *Schematic representation of the WES rationale and bioinformatic pipeline in CB HSPCs treated in vitro and retrieved from xenotransplanted mice in Fig. 19b.* ***b,*** *Venn diagrams representing variants sharing among in vitro treated samples from 'a'.* ***c-e,*** *Number of variants (c), number of SNV types (d) and their relative proportion (e)*

from in vitro sample from 'a' obtained upon subtraction of germline variants. ***f,*** *Reads alignments at B2M in the WES dataset from 'a'.* ***g-i,*** *Number of variants (g; median), relative proportion of variants (h; mean ± s.e.m) and relative proportion of SNV types (i; mean ± s.e.m) in the human xenograft from 'a' obtained upon subtraction of germline variants (n=3,2,3,4).* ***j,*** *Circos plots representing variants in cancer associated genes classified as high/moderate impact identified by WES in the human xenograft from 'a' (n=3,2,3,4).*

5.6 Optimized mRNA design improves efficiency and precision of base editing at target sites

To investigate whether the poorer performance and lower precision of BE4max could be improved by enhancing expression and lowering innate sensing, we engineered the mRNA constructs with a 5' cap better recapitulating the endogenous structure and included a eukaryotic Initiation Factor 4G (eIF4G) aptamer in the 5' UTR (**Fig. 22a**). Using these optimized mRNAs, we could decrease the effective mRNA dose and reach equivalent or superior editing efficiencies for all editing systems in both bulk and primitive HSPCs (**Fig. 22b,c**) nearly abolishing activation of IFN response (**Fig. 22d,e**) and lowering the p53 response across different target loci (**Fig. 22f,g**). When the optimized editors' mRNAs targeting *B2M* were co-delivered with an mRNA encoding for the p53 dominant negative mutant, GSE56, we abrogated p21 induction for all editors tested (**Fig. 22h**), albeit at the cost of slightly reduced efficiency and increased proportion of indels at the target site for BE4max (**Fig. 22i**).

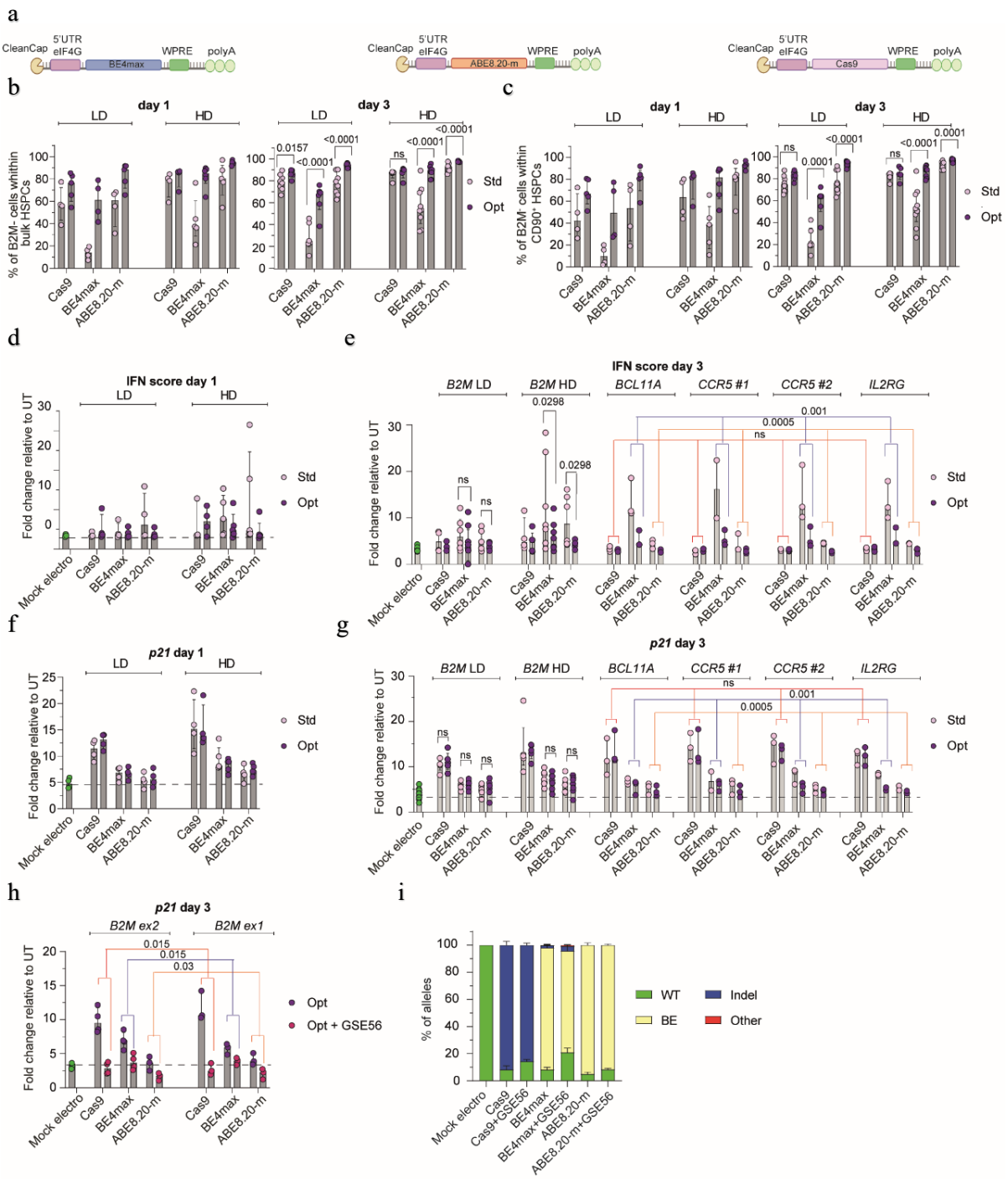


Figure 22. mRNA optimization improved editing efficiency and reduce IFN response. a, *Schematic representation of the optimized mRNAs.eiF4G: eukaryotic initiation factor 4 gamma.* ***b,c,*** *Percentage of B2M⁻ cells after editing at day 1 (left) or day 3 (right) post-thawing, measured by flow cytometry in bulk HSPCs (b) or in CD34+ CD133+ CD90+ HSPCs (c). Std/Opt: standard or optimized. LD: low dose (3.5 µg). HD: high dose (7.5 µg) (n=4,5,4,4, 4,5 for LD day 1; n= 4,4,5,6,5,5 for HD day1; n=8,9,6,6,8,9 for LD day 3; n=5,5,9,10,7,7 for HD day 3). Median with IQR. LME followed by post hoc analysis.* ***d,e,*** *IFN score defined as sum of fold change of IRF7, OAS1 and DDX58 expression over UT 24 hrs after editing at day 1*

(d) or day 3 (e) post-thawing (d: n=3 for Mock electro; n=3,4,4,4,4,4 for LD; n=3,4,5,5,5,5 for HD) (e: n=9 for Mock electro; n=4,5, 6,5, 5,5 for B2M LD; n=4,5, 7,7, 6,5 for B2M HD; n=3 for BCL11a, CCR5 and IL2RG). Median with IQR. For B2M LD and HD LME followed by post hoc analysis. For BCL11a, CCR5 and IL2RG Friedman test with Dunn's multiple comparison on unified samples. ***f,g,*** *Fold change of p21 expression over UT 24 h after editing at day 1 (f) or day 3 (g) post-thawing (f: n=4 for Mock electro; n=4 for LD; n=4,4,5,5,4,5 for HD); (g: n=11 for Mock electro; n=5,5,6,6,6,6 for B2M LD day 3; n=5,5,7,7,7,7 for B2M HD day 3; n=3 for BCL11a, CCR5 and IL2RG). Median with IQR. For B2M LD and HD LME followed by post hoc analysis. For BCL11A, CCR5 and IL2RG Friedman test with Dunn's multiple comparison on unified samples.* ***h,*** *Fold change of p21 expression over UT 24 hrs after editing at day 3 post-thawing with optimized mRNA in absence or presence of GSE56 (n=3). Median with IQR. Wilcoxon test on B2M exon 1 and exon 2 unified samples.* ***i,*** *Percentage of B2M exon 2 edited alleles, measured by deep sequencing, being WT or carrying the described editing outcomes (n=3). Mean ± s.e.m.*

On the contrary, indels induced at the target site by BE4max were significantly reduced when comparing optimized to standard mRNAs (**Fig. 23a,** and **Fig. 23b** to be compared with Fig. 14b since experiments were performed side by side with cells from the same HSPC donors). The apparently paradoxical decrease in indels while increasing the intended base editing by BE4max might be explained by higher co-expression of the UGI domains resulting in stronger inhibition of BER initiating factors. Consistently with this hypothesis, the proportion of indels was further lowered when editing at day 3 than at day 1, when BER associated genes are less expressed, and when using higher doses of mRNA. We obtained similar findings by screening 200 randomly picked HSPC-derived colonies for the occurrence of large deletions encompassing the *B2M* exon 2 target site (**Fig. 23c**). There were fewer colonies bearing large deletions from cells edited at day 3 with optimized BE4max mRNA as compared to standard BE4max mRNA, while Cas9 treated colonies showed, as expected, the opposite behavior (**Fig. 23c** to be compared with Fig. 15b since experiments were performed side by side with cells from the same HSPC donors). Similarly, translocations were not detected when using optimized BE4max mRNA while they were again found in Cas9 treated samples (**Fig. 23d** to be compared with Fig. 15f since experiments were performed side by side with cells from the same HSPC donors).

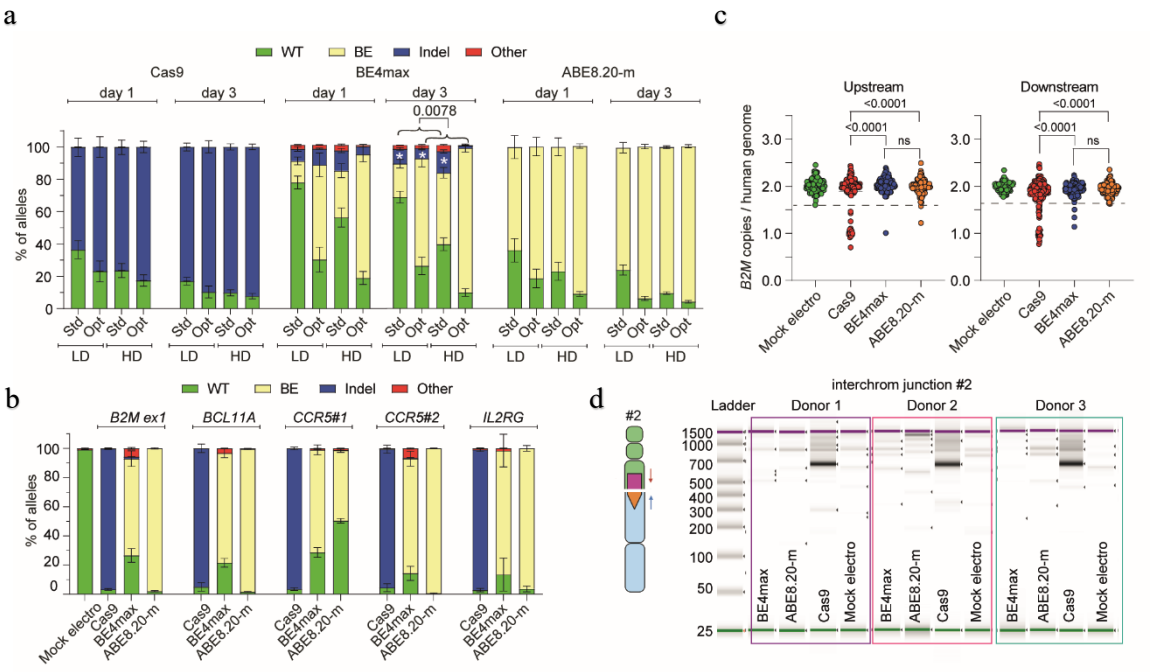


Figure 23. Optimized mRNAs reduced genotoxicity at the target locus. a, *Percentage of B2M exon 2 edited alleles, measured by deep sequencing, being WT or carrying the described editing outcomes (n=4). Mean ± s.e.m. Wilcoxon test performed on day 3 'Std' vs. 'Opt' groups unifying mRNA doses for statistical analysis.* ***b,*** *Percentage of B2M exon 1 (n=4), BCL11A, CCR5 and IL2RG (n=3) edited alleles, measured by deep sequencing, being WT or carrying the described editing outcomes. Mean ± s.e.m.* ***c,*** *Copies of B2M sequences per human genome flanking the target site in individual colonies generated by edited mPB HSPCs using optimized mRNAs (n=105,186,184,185 for 'upstream' assay; n=93,188,187,186 for 'downstream' assay). Dashed lines indicate the lower limit of the confidence interval from 'Mock electro' colonies. Median with IQR. Fisher's exact test.* ***d,*** *Images of capillary electropherogram showing amplification of #2 interchromosomal junction shown in Fig 15d upon HSPC editing with two gRNAs targeting B2M exon 2 and AAVS1 in 3 mPB donors.*

Furthermore, the use of optimized mRNAs at the lowest maximally effective dose allowed reaching >90% stable frequency of edited cells in the mouse xenografts for ABE8.20-m treatment and nearly 80% for BE4max (**Fig. 24a-d**). These levels of edited cells were also maintained in the human graft of secondary transplants (**Fig. 24e-h**). Indels at the edited site were low, albeit still detectable, in the human graft of primary recipients for both BEs, confirming that the optimized mRNAs not only increased efficiency but also precision of genetic outcome in long-term repopulating HSPCs (**Fig. 24i**).

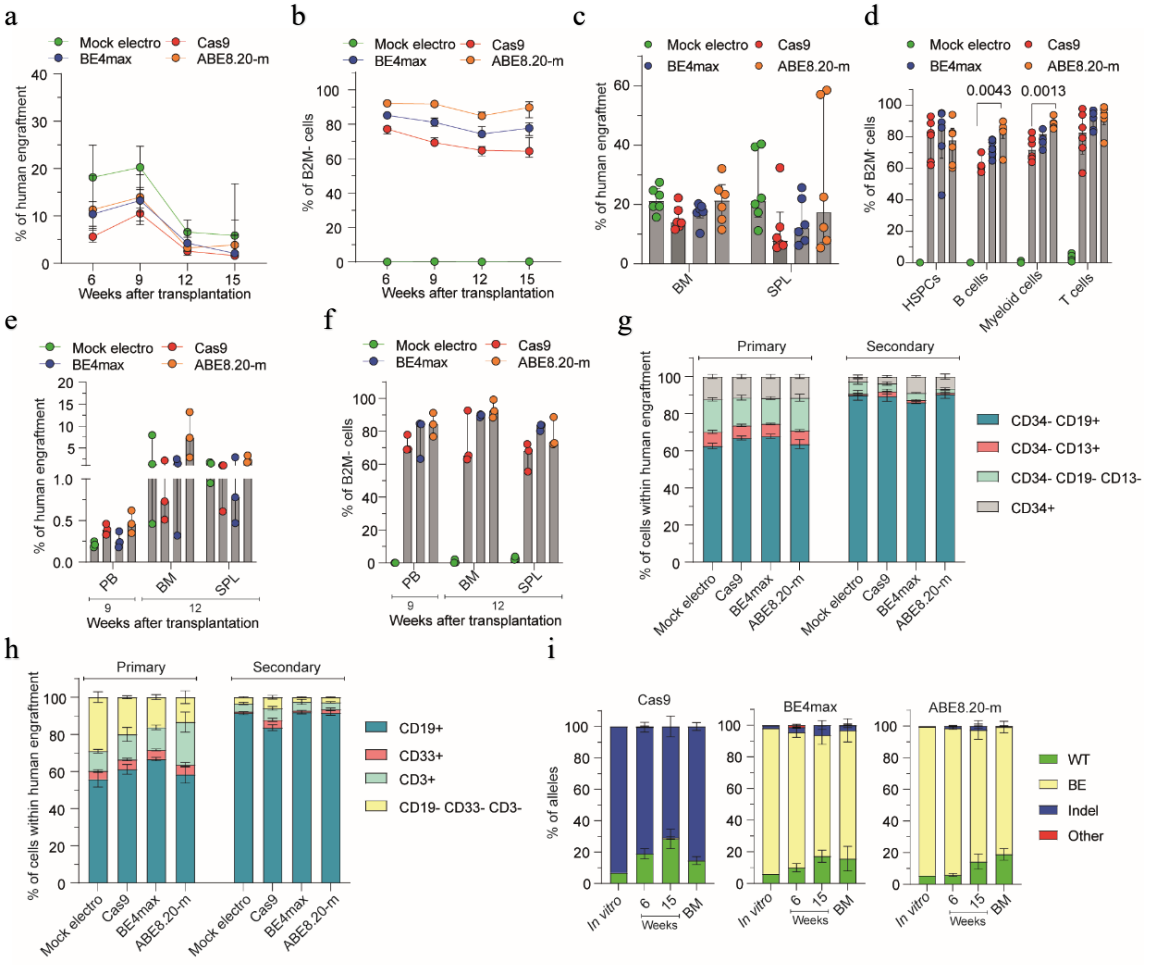


Figure 24. Optimized mRNAs improved editing efficiency upon xenotransplantation. ***a,b,*** *Percentage of human cells engraftment (a) and percentage of B2M⁻ cells within human graft (b) in mice transplanted with mPB HSPCs edited at day 3 post-thawing with optimized Cas9, BE4max and ABE8.20-m mRNAs at the lowest maximally effective dose (3.5, 7.5 and 3.5 µg, respectively) (n=6). Median with IQR. LME followed by post hoc analysis.* ***c,d,*** *Percentage of human cells engraftment (c) and B2M⁻ cells within hematopoietic lineages (d) in BM and SPL in mice form 'a' (n=6). Median with IQR. Kruskal-Wallis with Dunn's multiple comparison.* ***e,f,*** *Percentage of human cells engraftment (e) and B2M⁻ cells within human graft (f) in secondary recipient mice from 'a' (n=3). Median with range.* ***g,h,*** *Lineage composition in BM (g) and SPL (g) in mice from 'a' and 'e' (n=6,6,6,6,3,3,3,3). Mean ± s.e.m.* ***i,*** *Percentage of B2M exon 2 alleles, measured by deep sequencing, being WT or carrying the described editing outcomes in mice from 'a' and in the in vitro outgrowth (In vitro: n=1; Cas9: n=5; BE4max: n=6; ABE8.20-m: n=6). Mean ± s.e.m.*

5.7 Perturbation of exome mutational landscape emerges upon increased expression of BEs

We then evaluated whether the improved expression and activity of BE4max impacted the genome wide mutational landscape of treated cells and performed the same analyses described above in Fig. 21a (**Fig. 25a**). Differently from before, the total number of treatment-associated sequence variants was similar for optimized BE4max, standard ABE8.20-m and mock electro (**Fig. 25b**). Moreover, when we analyzed long-term engrafting clones, subtracting the donors' germline variants identified in the *in vitro* analysis, we found similar median numbers of variants among BE4max optimized and mock electro mice and a slight reduction in ABE8.20-m standard mice. No differences were retrieved in the relative proportions of SNV types among all transplanted mice and in the number of variants retrieved in cancer associated genes (**Fig. 25c-e**). These findings confirm a specific vulnerability of the BE4max editor, likely due to insufficient inhibition of the BER pathway, which results in loss of edited cells and oligoclonal grafts and is alleviated by improved expression of the editor. However, a concern remains that the impact of BE4max on the genomic mutational landscape emerged in the prior condition might now have escaped detection because of limited sensitivity in the context of more robust clonal abundance in the sample. We thus performed an orthogonal analysis on samples comprising a small, known and evenly distributed number of edited clones (**Fig. 25f**). We sequenced the exome of pools of 6 edit-bearing colonies outgrown from single-donor HSPCs treated with each different editor and expression construct and focused our bioinformatic analysis on mutations with a variant allele frequency compatible with the rate of *in vitro* accrual of mutations. Remarkably, this analysis confirmed previous findings of a slight increase but evident skewing of SNV types towards transversions for standard BE4max samples, as compared to mock, which was alleviated by the improved expression construct (**Fig. 25g-i**). Notably, in the latter optimized conditions, a trend towards increased proportions of the expected C>T/G>A transitions emerged over mock electro. On the other hand, cells treated with optimized ABE8.20-m construct showed an even higher increase in variants with skewed proportions towards transversions, cautioning that increased and/or prolonged activity of this type of editor might also increase adverse genome wide effects.

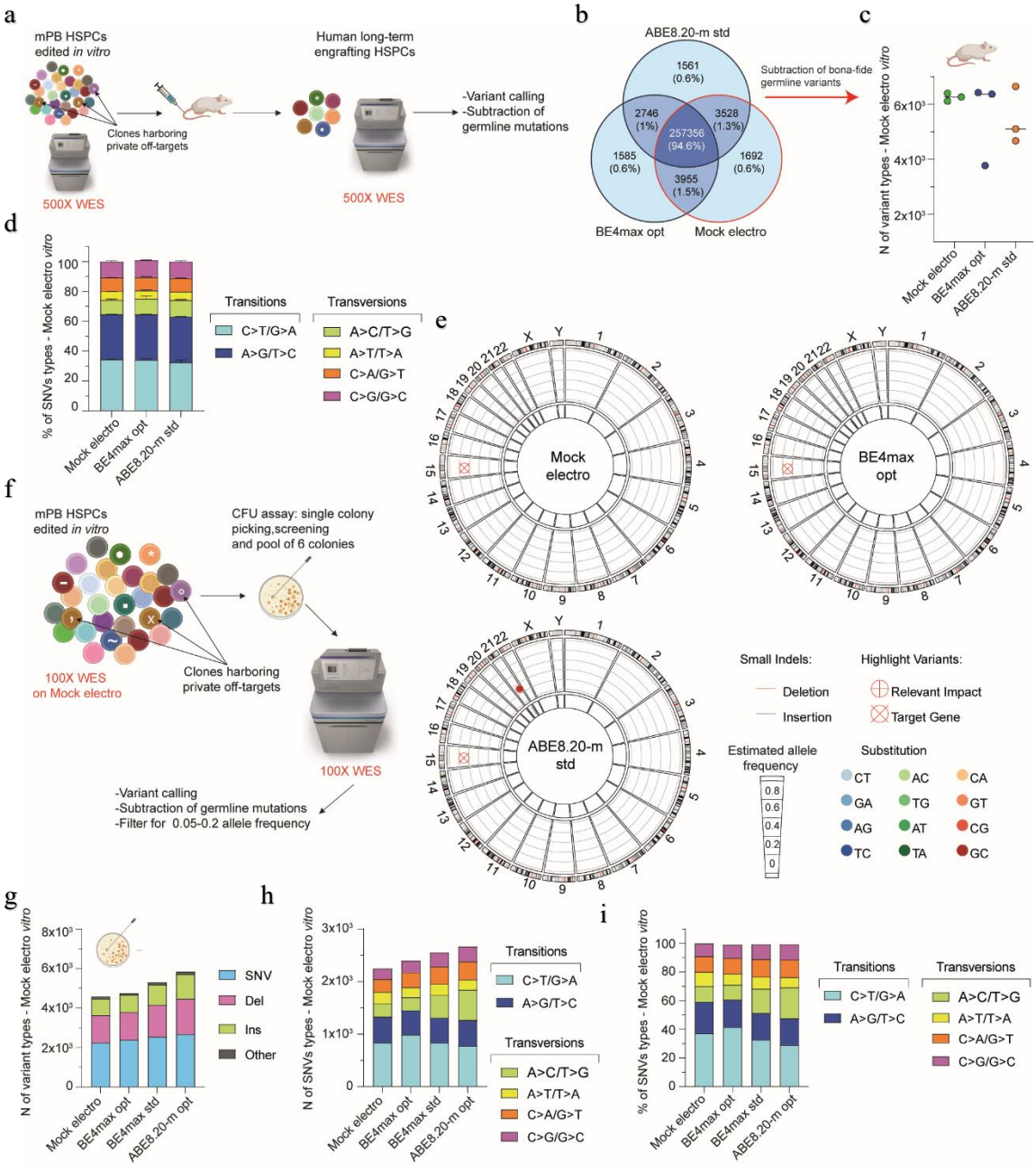


Figure 25. Mutational burden in exome of B2M edited HSPCs with optimized mRNA. ***a,*** *Schematic representation of the WES rationale and bioinformatic pipeline in mPB HSPCs treated in vitro and retrieved from xenotransplanted mice.* ***b,*** *Venn diagrams representing variants sharing among in vitro treated samples from 'a'.* ***c,d,*** *Number of variants (c; median) and relative proportion of SNV types (d; mean ± s.e.m)) in the human xenograft from 'a' obtained upon subtraction of germline variants.* ***e,*** *Circos plots representing variants in cancer associated genes classified as high/moderate impact identified by WES in the human xenograft from 'a'.* ***f,*** *Schematic representation of the WES rationale and bioinformatic pipeline in mPB HSPCs treated in vitro and retrieved from pool of individual colonies.* ***g-i,*** *Number of variants (g), number of SNV types (h) and their relative proportion (i) in the pool of colonies from 'f' obtained upon subtraction of germline variants.*

5.8 Efficient prime editing in human HSPCs does not escape DNA DSBs and cellular sensing

To broaden our investigation of nickase-based gene editors, we then included prime editing in our study. We first designed a panel of pegRNAs spanning *B2M* to induce its KO. Each pegRNA was also paired with a nicking gRNA to explore a PE3 setup, generating a nick on both DNA strands (**Fig. 26a**). No *B2M* modification was observed in K-562 cells for all pegRNAs tested except for pegRNA5, which induced 20% and 90% modified alleles when used without or with the cognate nicking gRNA, respectively (**Fig. 26b**). We then tested the selected PE3 setup in mPB-derived HSPCs from 6 independent healthy donors, treating cells after 3 days of culture, when expression of most genes belonging to DNA MMR pathway, which may antagonize prime editing, becomes lower (Ferreira da Silva *et al*, 2022; Chen *et al*, 2021d) (**Fig. 26c,d**). B2M KO cells were 30% in the bulk culture and 35% in the most primitive compartment (**Fig. 26e,f**) without detectable changes in composition of progenitor subsets (**Fig. 26g**). Molecular analysis revealed up to 60% modified *B2M* alleles, without detectable deletions spanning from one to the other nicking sites by PCR (**Fig. 26h,i**). Deep sequencing analysis of the *B2M* target site showed an average 40% precise prime editing outcome and 4.5% with additional insertion of the first bases of the pegRNA scaffold or small deletions at either nicking site (**Fig. 26j,k**).

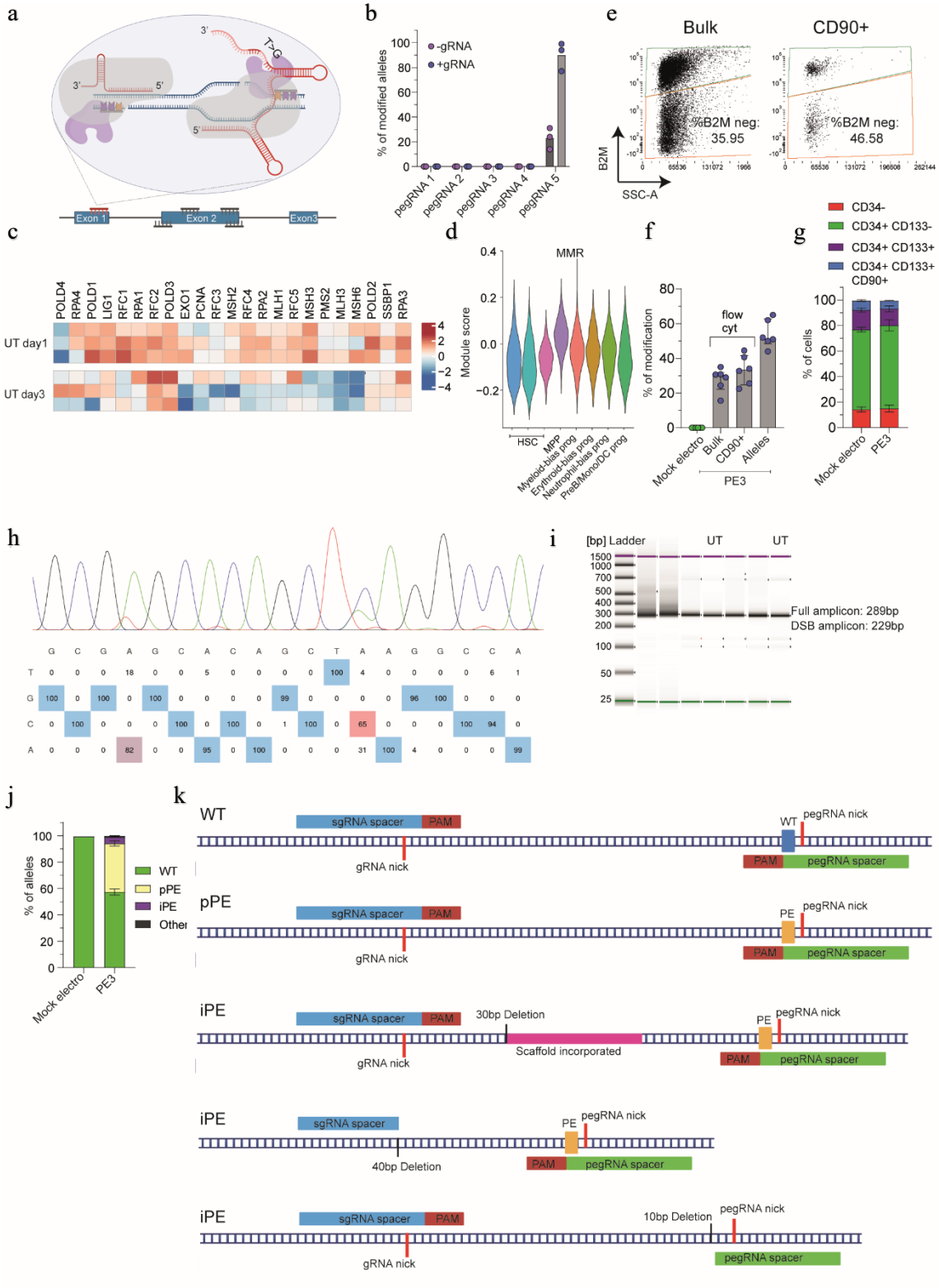


Figure 26. Efficient prime editing in mPB-derived HSPCs. a, *Schematic representation of the B2M prime editing screening. The selected pegRNA and gRNA are represented in red.* ***b,*** *Percentage of B2M prime edited alleles in K-562 cells measured by Sanger sequencing 9 days after editing procedure (n=3). Median.* ***c,*** *Heatmap of normalized read counts for genes belonging to the MMR pathway*

(KEGG database; hsa03430) in UT CB HSPCs cultured for 1 or 3 days. ***d,*** *Module score for genes belonging to the MMR pathway in different HSPC subsets from (Schiroli* et al*, 2019c).* ***e,*** *Flow cytometry plots showing the gating strategy for prime edited B2M⁻ mPB HSPCs.* ***f,*** *Flow cytometry (bulk and CD90+) and molecular analysis of B2M modification 7 days after prime editing in human mPB HSPCs (n=6). Median with IQR.* ***g,*** *Proportion of cellular subpopulations within mPB HSPCs from experiments in 'c' (n=6). Mean ± s.e.m.* ***h,*** *Representative plot of B2M Sanger sequencing.* ***i,*** *Representative image of capillary electropherogram of prime edited (n=5) and UT (n=2) samples.* ***j,*** *Percentage of B2M alleles, measured by deep sequencing, being WT or carrying precise prime editing (pPE), imprecise prime editing (iPE) or other modifications in mPB HSPCs (n=6). Mean ± s.e.m.* ***k,*** *Schematics of representative alleles for WT, pPE and iPE outcomes in prime edited samples from 'j'.*

Transcriptional analysis performed 24 hrs after PE3 treatment showed significant upregulation of genes related to IFN signaling (*IFI6* and *ISG15*), p53 activation (*CDKN1A* and *MDM2*) and unfolded protein response (*HSPA5* and *ATF3*) (**Fig. 27a**). Enrichment analysis confirmed activation of these pathways in PE3-treated cells, when compared to mock electro HSPCs (**Fig. 27b**). To stringently compare Cas9 and PE3 side by side, we combined Cas9 nuclease with *B2M* exon 1 gRNA that contains the spacer sequence of pegRNA5 as targeting region (**Fig. 27c**), and reached 80% and 50% allele modification, respectively, by the two systems (**Fig. 27d**). Cas9, but not PE3-treated HSPCs, showed a trend towards lower clonogenic capacity than mock electro cells, indicating a stronger impact of Cas9 than PE3 on HSPCs (**Fig. 27e**). Screening of around 140 randomly picked single colonies, revealed occurrence of large deletions after both treatments, although to lower extent with PE3 than Cas9 (**Fig 25f,g**). We next interrogated PE3- and Cas9-edited HSPCs for activation of selected IFN response and p53 pathway genes and found slight induction of the former ones in PE3 samples, and upregulation of the latter in both treatments, with lower extent for PE3 than Cas9, consistently with the above findings (**Fig. 27h-j**). We found selective activation of the pro-apoptotic isoform of *TP73* upon prime editing, which was completely absent when performing either Cas9 or base editing (**Fig. 27k-m**). Moreover, we found a mild but specific activation of *MT2A* in PE3 samples, supporting the activation of apoptotic responses (**Fig. 27n**). Overexpression of *TP73* by PE occurred in presence and absence of pegRNA in all 6 mPB HSPC donors tested (**Fig. 27o**) but was absent for a catalytically inactive RT fused with the nickase. However, a catalytically active RT fused with dCas9 failed to induce a similar response, showing that RT activity is necessary but not sufficient to induce the

proapoptotic *TP73* transcription. Altogether, these results suggest that *TP73* induction requires both RT activity and a concomitant nick at its DNA binding site, whether mediated by the nicking gRNA or pegRNA. Notably, the use of PE3max strategy, which improved prime editing efficiency by approximately 1.5-fold (**Fig. 27p**), prevented induction of *TP73* but not *MT2A* (**Fig. 27q**).

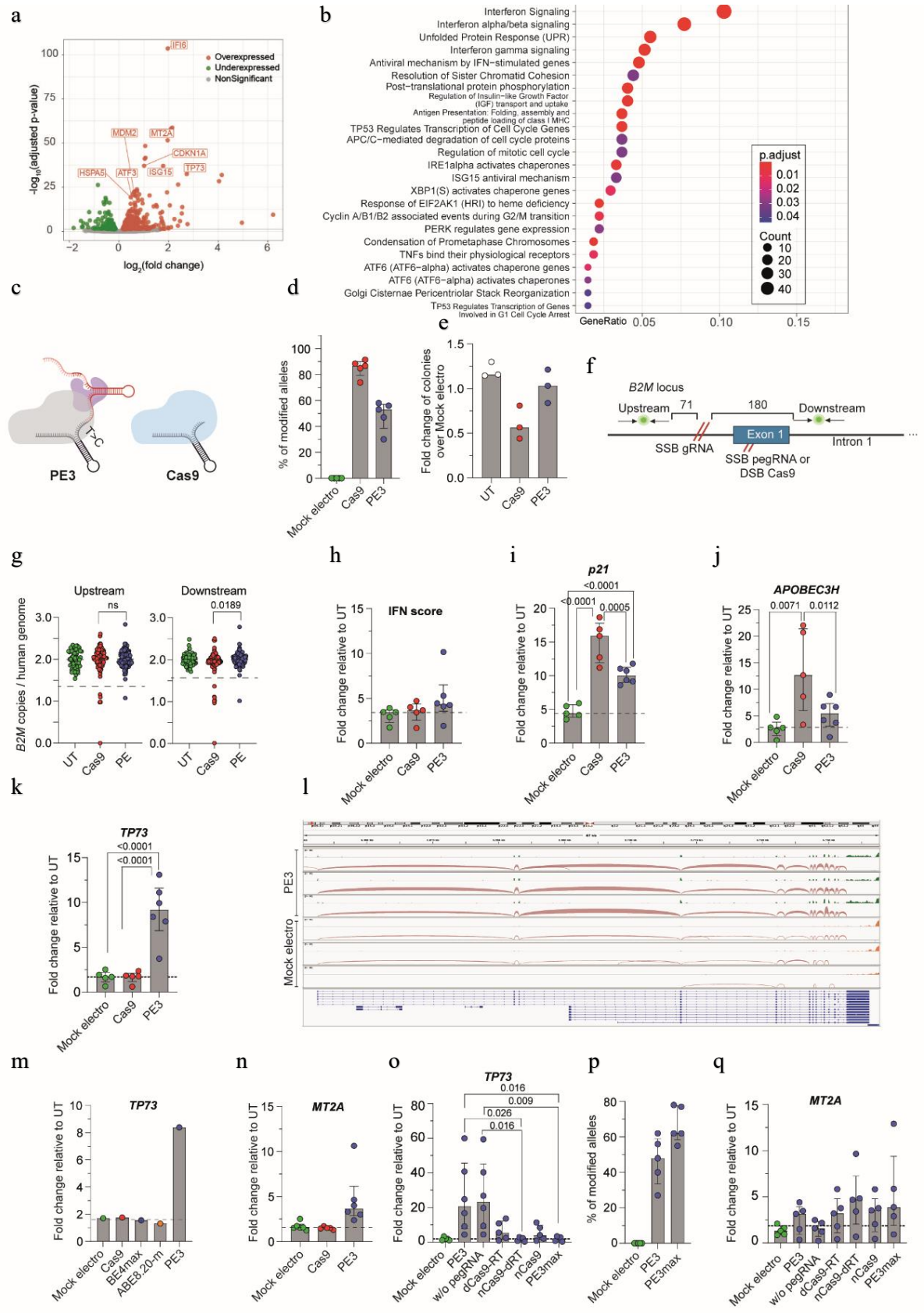

a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
Overexpressed
Underexpressed
NonSignificant
-log10(adjusted p-value)
log2(fold change)
IFI6
MDM2
MT2A
CDKN1A
HSPA5
ATF3
ISG15
TP73
Interferon Signaling
Interferon alpha/beta signaling
Unfolded Protein Response (UPR)
Interferon gamma signaling
Antiviral mechanism by IFN-stimulated genes
Resolution of Sister Chromatid Cohesion
Post-translational protein phosphorylation
Regulation of Insulin-like Growth Factor (IGF) transport and uptake
Antigen Presentation: Folding, assembly and peptide loading of class I MHC
TP53 Regulates Transcription of Cell Cycle Genes
APC/C-mediated degradation of cell cycle proteins
Regulation of mitotic cell cycle
IRE1alpha activates chaperones
ISG15 antiviral mechanism
XBP1(S) activates chaperone genes
Response of EIF2AK1 (HRI) to heme deficiency
Cyclin A/B1/B2 associated events during G2/M transition
PERK regulates gene expression
Condensation of Prometaphase Chromosomes
TNFs bind their physiological receptors
ATF6 (ATF6-alpha) activates chaperone genes
ATF6 (ATF6-alpha) activates chaperones
Golgi Cisternae Pericentriolar Stack Reorganization
TP53 Regulates Transcription of Genes Involved in G1 Cell Cycle Arrest
p.adjust
Count
GeneRatio
T>C
PE3
Cas9
% of modified alleles
Fold change of colonies over Mock electro
Mock electro
UT
B2M locus
Upstream
Downstream
Exon 1
Intron 1
SSB gRNA
SSB pegRNA or DSB Cas9
B2M copies / human genome
ns
0.0189
IFN score
p21
APOBEC3H
TP73
PE3
Mock electro
MT2A
BE4max
ABE8.20-m
w/o pegRNA
dCas9-RT
nCas9-dRT
nCas9
PE3max

Figure 27. Transcriptional response in prime edited HSPCs. a, *Volcano plot showing significant down- (green) and up- (red) regulated genes after prime editing.* ***b,*** *Dot plot of adjusted p-values of enriched categories on upregulated (FDR<0.05 and logFC >0) genes for PE3 versus mock electro HSPCs. Enrichment test.* ***c,*** *Schematic representation of the spacer sequence shared between pegRNA and gRNA for Cas9.* ***d,*** *Percentage of B2M Cas9 or PE3 edited alleles 7 days after treatments of mPB HSPCs (n=5). Median with IQR. Mann-Whitney test.* ***e,*** *Fold change in the number of colonies generated by mPB HPSCs over mock electro (n=3). Median.* ***f,*** *Schematic representation of the probes used for deletions detection. The distances between the target site and the closest primer of the ddPCR amplicons are shown.* ***g,*** *Copies of B2M sequences per human genome flanking the target site in individual colonies generated by edited mPB HSPCs (n=70,137,137 for 'upstream' assay; n=70,137,139 for 'downstream' assay). Dashed lines indicate the lower limit of the confidence interval from 'Mock electro' colonies. Median with IQR. Fisher's exact test.* ***h,*** *IFN score defined as sum of fold change of IRF7, OAS1 and DDX58 expression over UT 24 hrs after editing (n=5,5,6). Median with IQR. LME followed by post hoc analysis.* ***i-k,*** *Fold change of p21 (i) APOBECH3H (j) TP73 (k) expression over UT 24 hrs after editing (n=5,5,6). Median with IQR. LME followed by post hoc analysis.* ***l,*** *Representation of TP73 spliced alignment from RNA-Seq reads.* ***m,*** *Fold change of TP73 expression over UT 24 hrs after editing (n=1).* ***n,*** *Fold change of MT2A expression over UT 24 hrs after editing (n=5,5,6).* ***o,*** *Molecular analysis of B2M modification 7 days after prime editing in human mPB HSPCs (n=5). Median with IQR.* ***p,q,*** *Fold change of TP73 (p: n=5,6,6,5,5,5,5) and MT2A (q: n=5,6,6,5,5,5,5) expression over UT 24 hrs after editing. Median with IQR. LME followed by post hoc analysis.*

Prime edited HSPCs engrafted and persisted long-term in xenotransplanted mice maintaining >50% editing efficiency in PB and hematopoietic organs (**Fig. 28a-d**) The graft size of PE3 treated cells was reduced compared to mock electro cells, in particular at early times, conceivably due to a detrimental impact of the cellular responses described above on short-term repopulating progenitors. Deep sequencing analysis of the *B2M* target site on PB and BM cells revealed an average 42% precise and 5% imprecise prime editing outcomes (**Fig. 28e**). Prime edited HSPCs were able to engraft in secondary recipients, maintaining >50% efficiency in PB and hematopoietic organs (**Fig. 28f-i**) with no skewing of lineage compositions (**Fig. 28j,k**).

These data show that prime editing may reach substantial efficiency in long-term repopulating HSPCs and thus potentially broaden applications of genome editing to include transversion and other changes in the target sequence, with the current caveat of selecting an effective pegRNA. As also shown for BE, PE can still induce DNA DSBs and deletions at the target site, albeit to lower extent than nuclease-based editing and does

not escape cellular sensing of its unique machinery comprising nickase and reverse transcriptase.

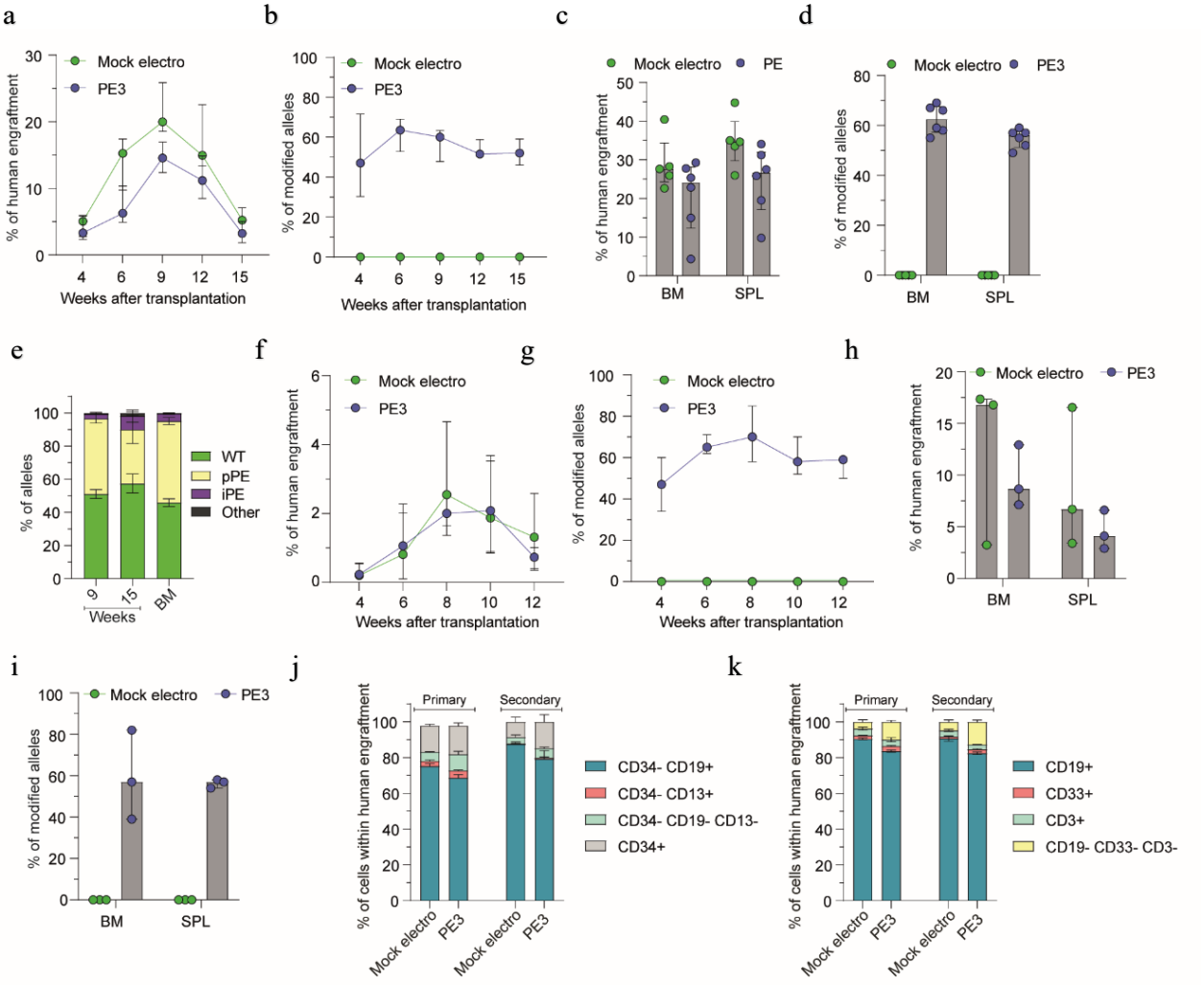


Figure 28. Assessing repopulation potential of prime edited HSPCs. a,b, *Percentage of human cells engraftment (a; LME followed by post hoc analysis) and percentage of B2M modified alleles within human graft (b) in mice transplanted with mPB HSPCs edited at B2M following PE3 or mock electroporation (n=5,6). Median with IQR.* ***c,d,*** *Percentage of human cells engraftment (c) and B2M modified alleles (d) in BM and SPL of mice from 'a' (n=5,6). Median with IQR.* ***e,*** *Percentage of B2M alleles, measured by deep sequencing, being WT or carrying pPE, iPE or other modifications in PB and BM of mice from 'a' (n=6). Mean ± s.e.m.* ***f,g,*** *Percentage of human cells engraftment (f) and percentage of B2M modified alleles within human graft (g) in secondary recipient mice from 'a' (n=3). Median with range.* ***h,i,*** *Percentage of human cells engraftment (h) and B2M modified alleles (i) in BM and SPL of secondary recipient mice from 's' (n=3). Median with range.* ***j,k,*** *Lineage composition in BM (p) and SPL (q) of mice from 'f' and 'i' (n=5,6,3,3). Mean ± s.e.m.*

5.9 Selection by means of artificial transactivators-2 (SMArT-2) allows enrichment of HDR-edited cells and purging of large deletions at the target site

As base and prime editing are constrained to addressing only single-point mutations or small insertions and deletions in humans HSPCs, Cas9-based HDR approaches remain the preferred option when aiming to paste a large therapeutic payload into a specific gene. However, as previously discussed in section 1.4.3 HDR is still limited in term of efficiency and precision in human HSPCs. Thus, we envisioned that an enrichment approach may broad the applicability of HDR gene editing.

To achieve transient expression of a selector gene upon HDR on-target integration, we design an AAV6 donor template carrying the corrective DNA sequence for *IL2RG* locus linked with the selector genes by means of self-cleaving peptides (2A) or internal ribosome entry sites (IRES). Upon on target integration the transgene is expressed under the control of the endogenous regulatory sequences of *IL2RG*. Overexpression of the selector (and the edited gene) can be achieved by delivering an artificial transactivator (ArT) (Adli, 2018) targeting the edited gene promoter, enabling enrichment of cells carrying targeted integration. Indeed, ArTs bind to the genome and work as transcription factor activating the expression of the downstream gene from a baseline level to robust overexpression. We called this strategy Selection by Means of Artificial Transactivator-2 (SMArT-2) (**Fig. 29**) which represents an evolution form SMArT-1 previously developed in the lab.

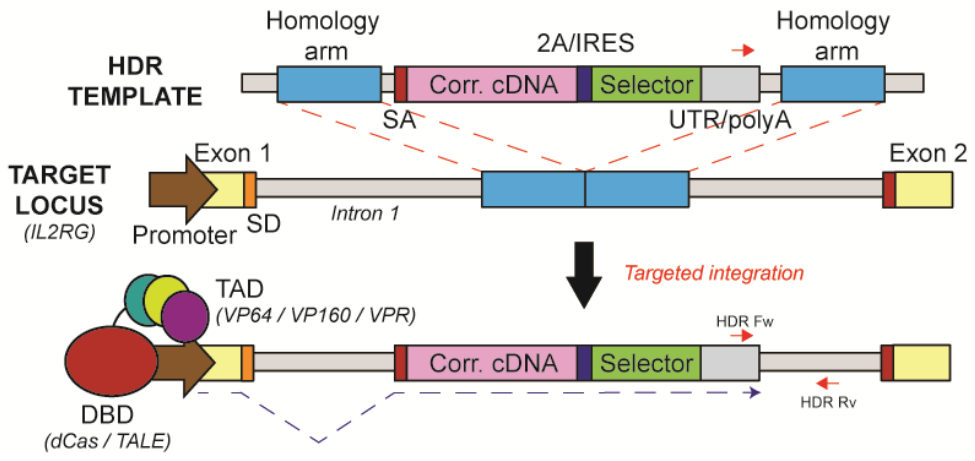


Figure 29. Schematic representation of SMArT-2 donor DNA and its on target integration in the AAVS1 locus. *Top: AAV6 donor DNA carrying the corrective cDNA fused with the selector. IL2RG target locus with complementary sequences to AAV6 homology arms. Bottom: representation of the integrated vector after HDR in which selector transactivation is regulated by ArT binding to the endogenous promoter region.*

As ArT, we screened a panel of orthogonal TALE DBD targeting the *IL2RG* promoter and fused each of them with a VP160 transactivating domain (i.e., 10 repeats of the Herpes Simplex Virus VP16 protein) (**Fig. 30a**). TALE3 ArT (T3) allowed 11-fold *IL2RG* overexpression compared to UT control in K-562 cell lines and was selected for further experiments. We then improved the ArT by coupling the T3 ArT with the VP64-p65-Rta (VPR) transactivator domain and delivered the construct as optimized mRNA in male mPB HSPCs. HSPCs were electroporated with RNP targeting *IL2RG* and optimized mRNAs for the T3-VPR and the p53 inhibitor GSE56. After 15 minutes, cells were transduced with an AAV6 vector carrying the corrective IL2RG cDNA in frame with dNGFR as selector. While in absence of ArT delivery, dNGFR expression was low in bulk HSPCs and nearly absent in the most primitive CD90+ compartment, the co-delivery of the optimized T3-VPR allowed robust expression of the dNGFR selector (**Fig. 30b,c**). Molecular analyses by ddPCR showed that the selection efficiency, measured as fraction of HDR-edited HSPCs also expressing dNGFR, averaged at 75% (**Fig. 30d**). We next investigated at clonal level the on-target genetic configuration of bulk and sorted HSPCs by screening hundreds of single HSPC colonies with ddPCR assays probing for HDR and long-range deletions. SMArT-2 enriched at 97% efficiency HSPCs edited at 55% HDR efficiency in the bulk population, confirming the possibility to enrich for HDR edited cells (**Fig. 30e**). Concordantly, HDR-edited HSPCs were partially depleted in colonies originating from selector-negative HSPCs. Long-range deletions encompassing the *IL2RG* promoter were absent in colonies seeded from selector-positive HSPCs, while found at 1.6 and 3.0% frequency in colonies originating from bulk and selector-negative HSPCs, respectively (**Fig. 30f**). This data highlights the potential of SMArT-2 in enriching for HDR-edited HSPCs while purging for those bearing unintended and potentially genotoxic on-target editing outcomes.

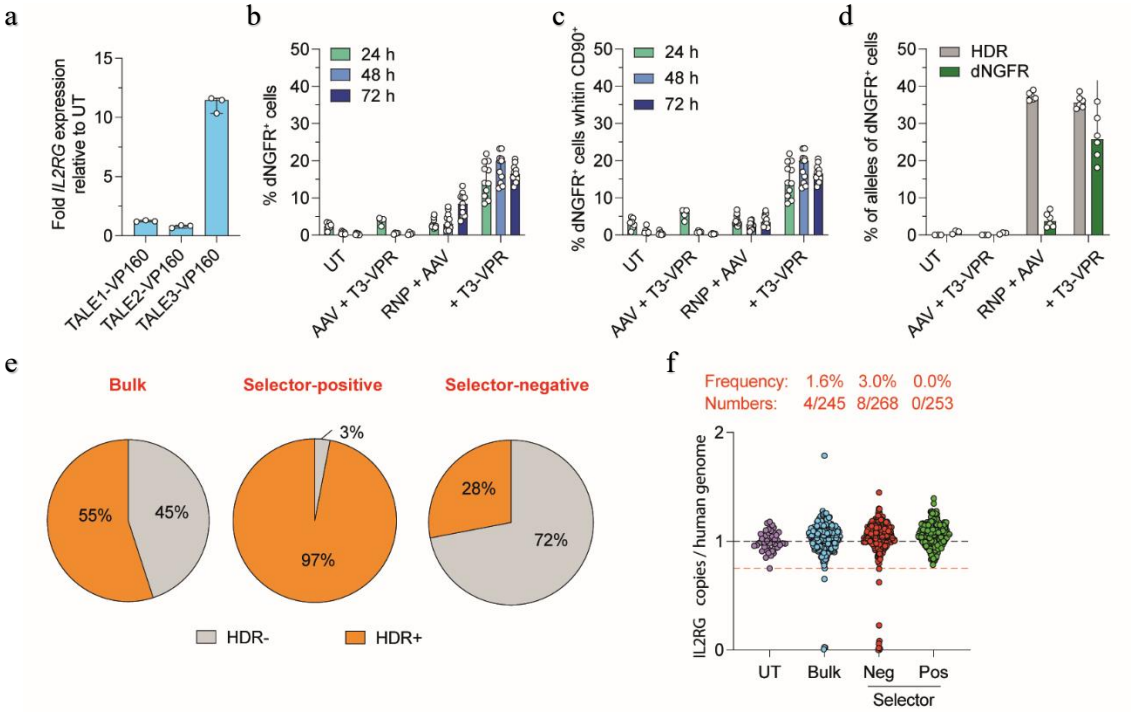


Figure 30. SMArT-2 allows enrichment of HDR edited HSPCs. a, *IL2RG fold expression over UT upon SMArT-2 donor integration and transactivation.* ***b,c*** *Percentage of dNGFR positive bulk (b) or CD90^{+} (c) HSPCs.* ***d,*** *Percentage of HDR edited alleles as measured by ddPCR and Percentage of NGFR+ cells as measured by FACS.* ***e,*** *Percentage of HDR in bulk, selector positive or selector negative HSPCs.* ***f,*** *Copies of IL2RG sequences per human genome upstream the target site in individual colonies generated by SMArT-2 edited mPB HSPCs from bulk of from sorted negative or positive fractions.*

5.10 SMArT-3 strategies allows enrichment of HDR-edited HSPCs

To prevent selector co-expression with the endogenous gene which may lead to toxicity and immunogenicity in some contexts, we designed an alternative strategy called SMArT-3. We designed an AAV6 donor template targeting the genomic safe harbor *AAVS1* where sustained therapeutic transgene expression can be achieved without perturbing the neighbor gene regulation and the epigenetic landscape (Lombardo *et al*, 2011)(**Fig. 31a**). The AAV6 donor contains a minimal CMV promoter, comprising only the TATA box (minCMV) driving the expression of the selector gene (dNGFR or GFP), which can be coupled with an independent cassette for the gene of interest. To obtain transient expression of the selector restricted to cells that experienced HDR, we designed ArTs binding to the genome sequence flanking the homology arm for HDR.

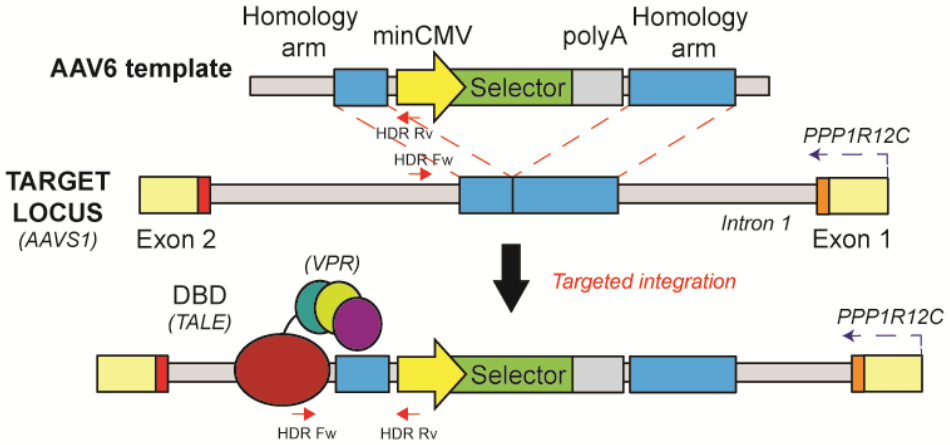


Figure 31. Schematic representation of SMArT-3 donor DNA and its on target integration in the AAVS1 locus. *Top: AAV6 donor DNA carrying selector cassette and gene of interest. Middle: AAVS1 target locus with complementary sequences to AAV6 homology arms. Bottom: representation of the integrated vector after HDR in which selector transactivation is regulated by ArT binding to the genomic DNA upstream of the selector cassette.*

We screened a panel of TALE DBD fused with the VPR transactivator by delivering them as plasmids in a K-562 cell line clone (H1), stably carrying precise monoallelic HDR-mediated on-target integration of a minCMV promoter sequence regulating dNGFR selector expression, which was undetectable on cell surface at steady state (**Fig. 32a,b**). TALE#7 (T7)-VPR resulted in the highest selector transactivation and was selected for further studies. We next tested SMArT-3 in mPB HSPCs in which we co-delivered together with the optimized editing reagents (including RNP targeting *AAVS1* locus and mRNA encoding for GSE56/E4orf6/7) an optimized mRNA encoding for the T7-VPR. mPB-HSPCs were then transduced with an AAV6 containing minCMV promoter and GFP as selector. Optimized T7-VPR mediated robust GFP expression in a median of 35% bulk HSPCs and 25% primitive CD90+ HSPCs 48-72 hours after editing. By contrast, selector expression was not detectable in absence of ArT (**Fig. 32c,d**). Of note, relative fluorescence intensity (RFI) of GFP^+ against GFP^- cells was half in the most primitive compartment than in the bulk HSPC population (**Fig. e-g**), suggesting lower permissiveness of this primitive cellular subset to ArT delivery and/or activity. The percentage of *AAVS1* HDR-edited alleles in the bulk population was 45-50% with a trend for slight reduction in the T7-VPR conditions (**Fig. 32h**), possibly due to competition either between the DNA repair machinery and T7-VPR DNA binding or between T7-VPR and GSE56/E4orf6/7 mRNAs for protein translation. Importantly, FACS-based

enrichment of GFP+ cells significantly increased the fraction of HDR-edited alleles in the GFP+ fraction up to 80%, providing the proof-of-concept of selection of HDR-edited cells by SMArT-3 (**Fig. 32i**). However, persistence of a considerable fraction of HDR-edited alleles in the sorted GFP- fraction suggested high stringency of the SMArT-3 strategy at the expense of the selection yield. To evaluate the genotype composition and integrity of the target site, we analyzed single colonies in sorted and bulk edited HSPCs and found that 97% of sorted GFP+ colonies carried at least one HDR-edited allele (**Fig. 32j**). Moreover, we found a trend towards purging of clones carrying long-range deletion spanning over the T7 DNA binding site in the GFP+ fraction (**Fig. 32k**).

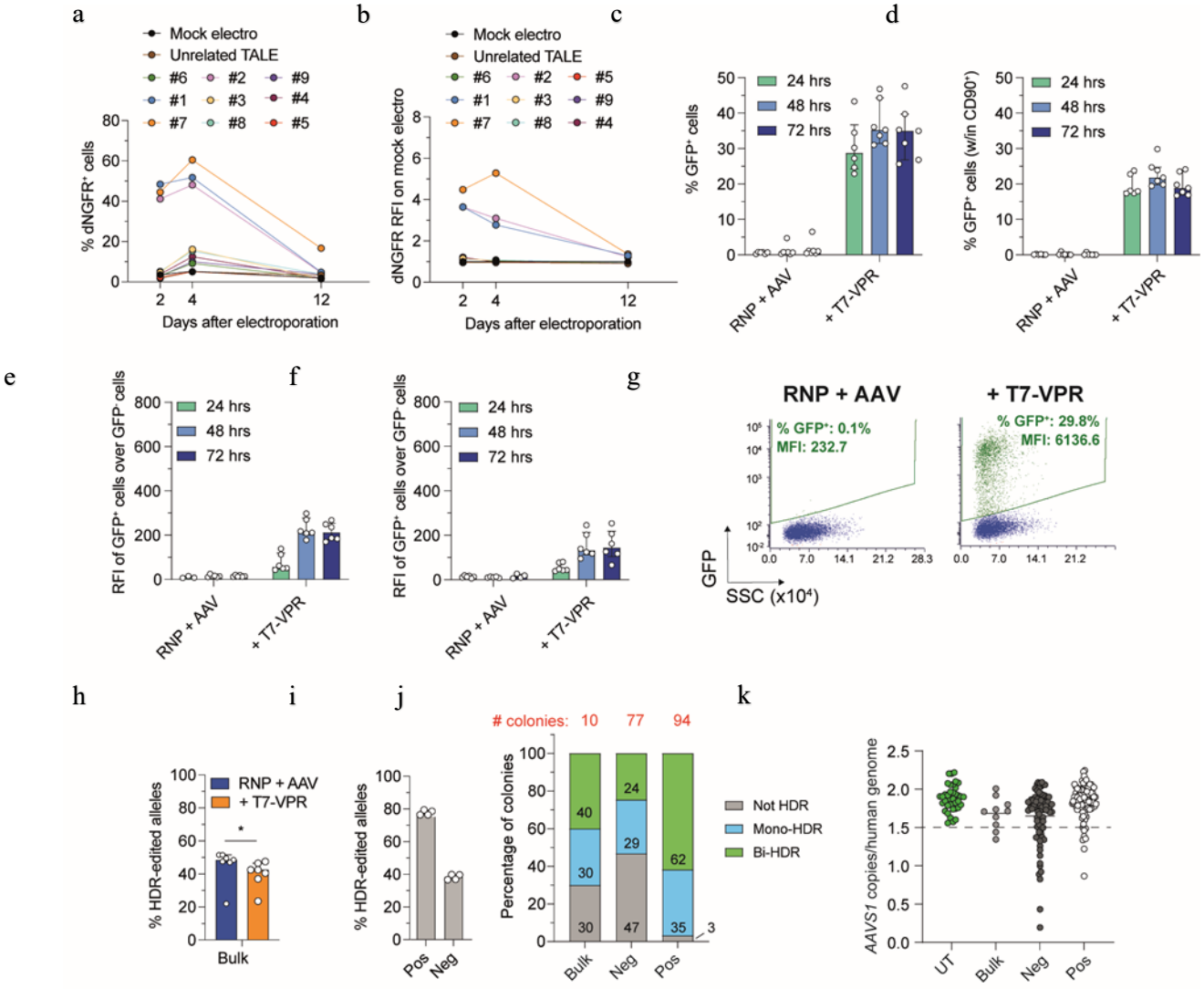


Figure 32. Successful in vitro selection of SMArT-3 edited HSPCs. a,b, *Percentage (a) and relative fluorescence intensity over untreated control (b) of NGFR+ cells at 2,4 and 14 days after electroporation of different sgRNAs and ArT in a K562 cell line clone harboring on-target integration of the minCMV-NGFR selector cassette (n=1).* ***c,d,*** *Percentage of GFP⁺ cells in bulk (c) or CD90⁺ (d) mPB*

derived HSPCs upon HDR editing in presence (-T7-VPR) or not of the ArT (n=6,7,7) Median with IQR. ***e,f,*** *Relative fluorescence intensity over GFP$^-$ control of bulk (e) or CD90+ (f) mPB derived HSPCs upon HDR editing in presence (-T7-VPR) or not of the ArT (n=6,7,7) Median with IQR.* ***g,*** *Rapresentative plot of bulk mPB-HSPCs HDR edited in presence (-T7-VPR) or not of the ArT.* ***h,i*** *Percentage of HDR edited alleles measure by ddPCR in bulk HSPCs (h) or in sorted GFP+ and GFP- fractions (i) (h: n=7; i: n=4), Median with IQR.* ***j,*** *Percentage of unedited (not HDR), monoallelic and biallelic edited individual colonies plated from bulk edited mPB-HSPCs and in sorted GFP+ and GFP- fractions. (n=10, 77, 94). Mean ± s.e.m.* ***k,*** *Copies of AAVS1 sequences per human genome upstream the target site in individual colonies generated by SMArT-3 edited mPB HSPCs from bulk of from sorted GFP+ and GFP- fractions.*

To assess the repopulation potential of SMArT-3 enriched HSPCs, we transplanted an equivalent number of bulk-edited, sorted GFP+ or GFP- HSPCs from a pool of human donors in immunodeficient mice. Transplantation of sorted GFP+ HSPCs resulted in 5- to 10-fold lower human cell engraftment than bulk and sorted GFP- counterparts in both peripheral blood and hematopoietic organs at the end of the study (**Fig. 33a,b**). This finding is in line with the lower HDR efficiency and GFP marking expected in the long-term engrafting HSPCs compared to more committed progenitors, leading to overrepresentation of differentiated cells in the selected and infused cell product. Strikingly, HDR efficiency in the sorted GFP+ group was 5-fold higher than in the bulk one, ranging between 1 and 1.5 average copies of edited alleles per cell (**Fig. 33c,d**), suggesting the achievement of a fully edited graft. Importantly, GFP expression was undetectable in all the groups, confirming the lack of basal selector expression (**Fig. 33e**).

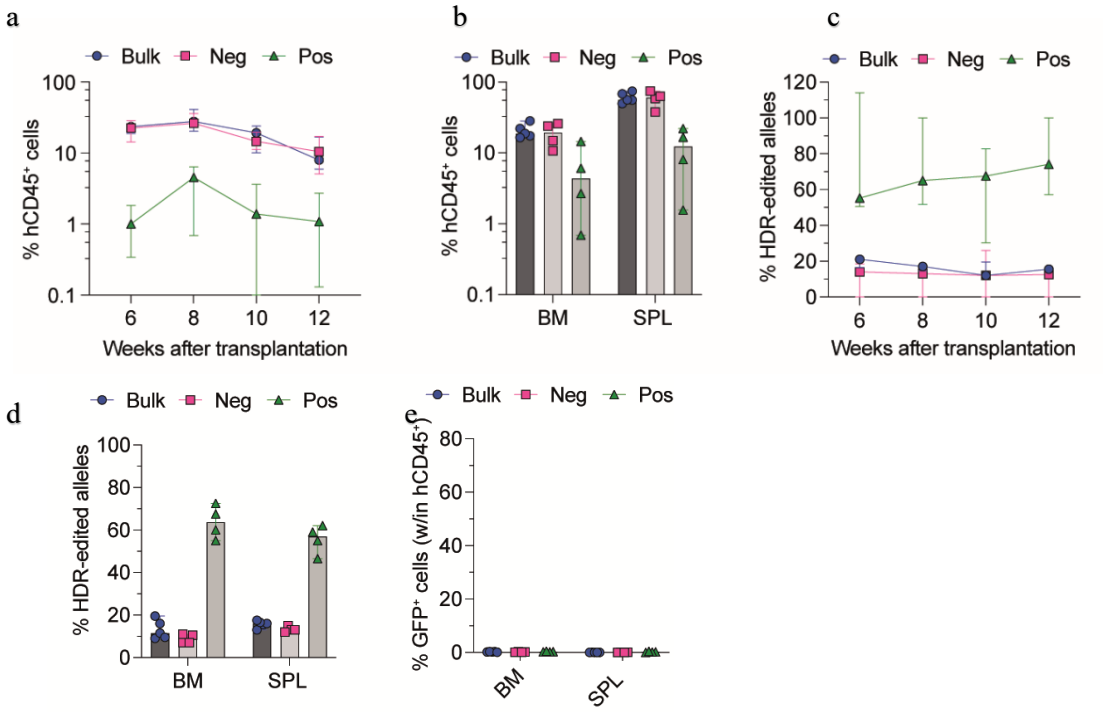


Figure 33. In vivo persistence of SMArT-3 edited HSPCs. a,b, *Percentage of human cells engraftment in PB (a) and hematopoietic organs (b) in mice transplanted with bulk edited mPB HSPCs or sorted sorted GFP+ and GFP- fractions. (n=5,4,4). Median with IQR.* ***c,d****, Percentage of HDR edited alleles in PB (c) and hematopoietic organs (d) in mice from 'a'. (n=5,4,4). Median with IQR.* ***e,*** *Percentage of GFP expression in BM and SPL of mice from 'a'.*

Finally, we tested the therapeutic configuration of the SMArT-3 strategy by including in the AAV6 donor template an independent cassette driven by a constitutive promoter. We substituted minCMV promoter with its improved version 'T6-minCMV' (SK), (Loew *et al*, 2010) with lower basal transcriptional activity, to drive selector expression (dNGFR) and included a PGK promoter to drive the expression of the gene of interest (GOI) (GFP) (Fig. 34a). Flow cytometry data on mPB-derived HSPCs treated with SMArT-3 strategy revealed constitutive expression of the GFP (Fig.34b), while selective dNGFR expression only in presence of T7-VPR confirming the feasibility selection of cells harbouring on target integration of GOI expressed at supraphysiological levels in a safe harbour locus (Fig. 34c). Overall, these data demonstrate the feasibility of an enrichment approach to select long term engrafting HDR-edited HSPCs and highlight the potential of SMArT-3 to reduce carryover of cells with unintended edits at the target site.

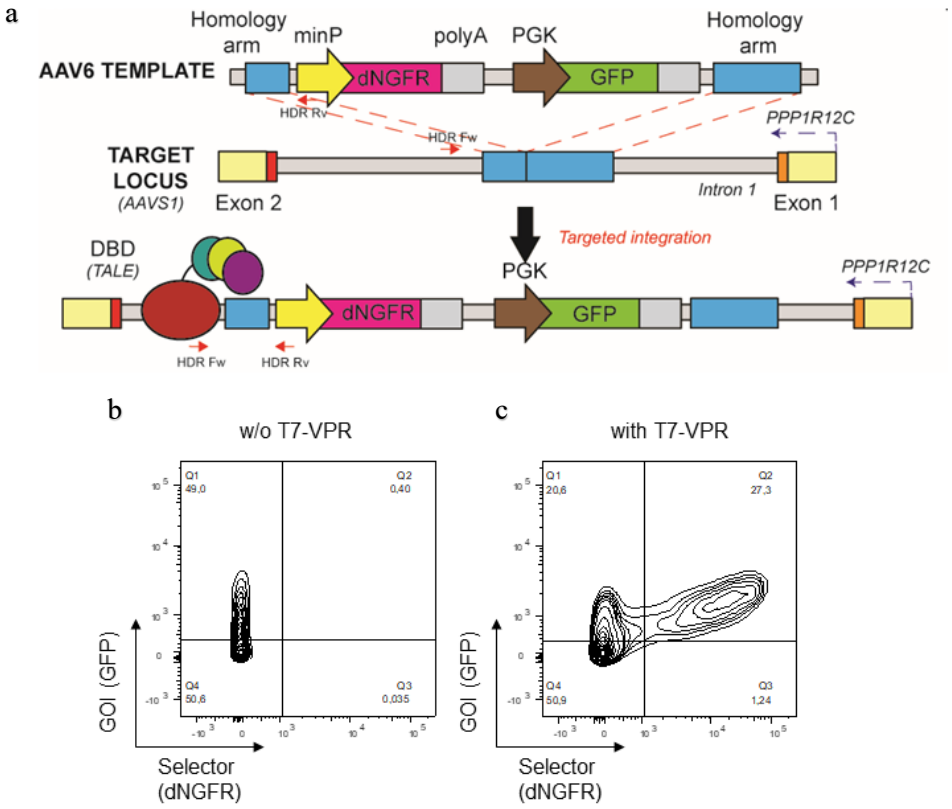


Figure 34. Therapeutic configuration of SMArT-3 strategy. ***a,*** *schematic representation of the SMArT-3 and its repair outcome.* ***b,c,*** *flow cytometry plot of mBP derived HSPCs transuced with SMArT-3 donor template in absence 'b' or presence 'c' of T7-VPR.*

6 DISCUSSION

In this study, we investigated the application of state-of-the-art nickase-based editing systems, including CBE, ABE, and PE, on human HSPCs, comparing them to conventional Cas9 nuclease-based editing. Our findings elucidated that CBE, ABE, and PE induce DNA DSB at the target site, although to a lesser extent compared to Cas9. This raised concerns about potential genotoxic byproducts such as translocations and large deletions. Notably, the extent of these events was more pronounced with CBE, primarily due to the concurrent activity of BER, but this effect was mitigated through the use of an optimized mRNA construct, resulting in improved expression and complete BER inhibition. Furthermore, our research revealed detrimental transcriptional responses across all platforms, which impaired editing efficiency and the repopulation capacity of HSPCs. These effects were partially alleviated through editor engineering. Finally, WES analyses revealed that exposure to base editors induced changes in the mutational landscape of edited HSPCs. These changes are challenging to capture, and their impact remains unpredictable.

In parallel, in the context of HDR-based applications, we developed novel strategies for the enrichment of HDR-edited cells, which allow to increase the proportion of engrafting HSPCs carrying the intended genetic modification and to purge our those carrying imprecise, non-corrective on-target edits.

6.1 Side-by-side comparison of cutting-edge editing technologies in terms of efficiency and precision at the target site.

The differences in efficiency observed for nuclease vs. nickase editing and among the different BEs and PE, even when targeted to the same locus, likely reflect the different biochemical reactions captured to install the edits and the type of genome configuration and cellular environment where the source enzyme naturally evolved. Indeed, the higher efficiency of ABE may reflect the absence of an endogenous antagonizing pathway in human cells, which instead is present and hinders fixation of the intended mutation both for CBE and PE. Furthermore, we found that, as compared to conventional Cas9 nuclease, CBE and ABE as well as PE lower but do not abolish occurrence of DNA DSBs at their genomic targeted sites, exposing the cells to the potential genotoxic effects of deletions and translocations. It is conceivable that intrinsic features of the locus may influence the

efficiency and precision of base and prime editing. For instance, the presence of multiple editable nucleotides in the target sequence (as in the case of BE4max for the *B2M* exon 2 target site) may cause tandem deaminations and affect the type and kinetics of repair, leading to different proportion of byproducts in the outcome. Moreover, bystander editing may limit the use of base editing when aiming to correct specific disease-causing mutations. Similarly, the challenge in designing an efficient pegRNA, as reported here, highlights a potential hurdle of the current PE system that may be overcome by next-generation molecules and trained algorithms (Hsu *et al*, 2021; Mathis *et al*, 2023). Concerning the mechanisms underlying the generation of DNA DSBs at the target site, there may be common and specific factors for each editing system. Conversion of a SSB to DSB upon transit of a DNA replication fork is likely a shared mechanism among all nickase-based systems (Kuzminov, 2001). The higher fraction of alleles carrying deletions and the stronger propensity to generate translocations observed for the CBE likely reflects the involvement of BER by UG recruitment at uracil nucleotide residues and subsequent APEX1-dependent nick. If this repair process is not inhibited, the combination of APEX1 and BE-dependent nicks on the two opposite strands may result in a staggered DNA DSB at the target site, which may be repaired eventually by NH-/MM-EJ. Supporting this explanation is the observation of larger proportions of alleles carrying deletions in BE4max-treated cells at day 1 compared to day 3, when the BER machinery is upregulated. Similarly, the decrease of alleles carrying indels over time in the graft, when the output of long-term repopulating cells become prevalent, may correlate with the lower BER gene expression in primitive vs. committed progenitors, a finding concordant with a previous report on murine HSPCs (Beerman *et al*, 2014). Furthermore, we showed that while more robust expression of Cas9 by mRNA optimization increased the proportion of indels at the target, we observed the opposite trend for BE4max, with fewer deletions and more stable and polyclonal edited graft *in vivo*. This apparent paradox might well be explained by more robust inhibition of UG activity by the UGI domains coupled to the BE. On the other hand, it is possible that the higher precision of the ABE vs. CBE may reflect a lower processivity and binding affinity of the former editor for eukaryotic DNA as compared to its natural activity on bacterial RNA (Neugebauer *et al*, 2023). Even though the frequencies of deletions and translocations were lower upon nickase than nuclease-based editing, these figures remain

relevant considering that several hundred million of HSPCs are treated and infused in clinical applications of HSPC gene therapy (>2×106 CD34+ cell/kg) (Canarutto *et al*, 2021). Hence, the potential occurrence and *in vivo* persistence of large genomic rearrangements should be considered in the preclinical and clinical risk assessment of base and prime editing and, even more stringently, when aiming for multiplex editing approaches. In the latter context, epigenome editing might eventually provide an intriguing alternative for targeted gene knockout (Nuñez *et al*, 2021; Amabile *et al*, 2016).

6.2 Dissecting and overcoming transcriptional responses which impair repopulation potential of edited HSPCs

In our study, we uncovered that all systems induce detrimental transcriptional responses in the treated cells that negatively impacted editing efficiency and/or clonogenic and repopulation capacity. While some of the responses were common between Cas9 and cutting-edge platforms, we also uncovered specific responses to BE and PE which were previously unappreciated. Activation of IFN response was observed upon delivery of long and complex mRNA structures and may contribute to lower engraftment of edited cells, in particular for long-term repopulating progenitors. This hurdle was nearly completely overcome by optimizing RNA design to increase yield, purity and stability. For instance, the delivery of editor protein instead of mRNA might avoid responses due to mRNA sensing and possibly mitigate unwanted editing outcomes, despite their production and purification remains challenging to afford and standardize. Furthermore, as previously reported, p53 pathway activation consequent to DNA DSB strongly impacted the size and clonality of the human graft in transplanted mice (Ferrari *et al*, 2020a). While this response was well evident for Cas9 nuclease editing, both base and prime editing induced detectable activation of p53 target genes, with PE and BE4max being higher than ABE8.20-m and resulting in lowered engraftment compared to control. The induction of p53 target genes observed for each system correlated to some extent with the proportion of indels and large deletions found at the target site, formally proving induction of DNA DSB. The activation of described pathways, particularly p53, affect the clonal distribution in the graft, as also shown here for Cas9-treated HSPCs. Conversely, our clonal tracking analysis did not detect decrease number of repopulating clones nor aberrant expansion following BE, suggesting that the induced insults do not alter the overall HSPC fitness. Since this analysis was performed on grafts with different

proportion of edited cells, i.e., 100% and 30% biallelic-edited respectively for ABE and CBE, the impact of the latter might be partially masked. Interestingly, DNA damage response was abrogated in the context of base editing when transiently inhibiting p53, albeit at the cost of slightly reduced efficiency for all systems, possibly due to competition between editors and p53 inhibitor mRNAs' entry and translation. Yet, the increased proportion of indels at the CBE target site suggested reduced purging of clones experiencing higher DNA damage burden and discourages from adopting p53 inhibition in this context. Regarding prime editing, PE3 showed specific induction of the pro-apoptotic transcript of *TP73* which might be consequence of the formation of on-target editing intermediates, induced by the concurrent DNA nick and local RT, especially when not rapidly resolved, and/or of MMR activation. These hypotheses are respectively in line with the reported *TP73* induction upon pharmacological topoisomerase inhibition in eukaryotic cells (Al-Bahlani *et al*, 2011) and with *TP73*-dependent apoptosis triggered by MMR (Long *et al*, 2008). While we uncovered how detrimental responses impairs the engraftment potential of prime edited HSPCs compared to mock electro control, a clonal tracking analysis will help in dissecting their impact on the clonal complexity and dynamics during repopulation. Whereas most of our findings were conserved across multiple target sites and HSPC sources, BE and PE are constantly being evolved (Anzalone *et al*, 2020), leading to superior efficiencies, sometimes at the cost of targeting specificity. Continuous engineering guided by the type of experimental findings reported here as well as the rationale design of next generation editing systems may allow further performance improvement of these transformative tools (Lam *et al*, 2023; Neugebauer *et al*, 2023).

6.3 Base editing results in gRNA-independent off-target activity on the genome

One of the most challenging aspects of investigating the specificity of emerging editing systems combining a nCas9 domain for tethering the editor to the intended target with a constitutively active enzyme, is the possibility of gRNA-independent global activity of the latter on the transcriptome and on the genome. Despite off-target deamination in the transcriptome are generally considered less dangerous compared to off-target deamination in the genome since they do not result in permanent and heritable changes, they still can transiently affect gene expression and cellular function. Regarding off-target

in the genome, they may escape detection when interrogating complex mixtures of treated cells, as bulk *in vitro* cultures or highly polyclonal grafts, because of dilution and lack of recurrence in the experimental context. On the other hand, analysis of samples comprising the expanded outgrowth of a known or predicted small number of clones might help uncovering an altered frequency or distribution of variants associated with specific treatments, as shown here for the clonally shrunken graft of BE4max edited cells or pools of *in vitro* colonies formed by edited cells. Moreover, the engagement of different DNA repair pathways and genomic surveillance mechanisms by multiple concurrent DNA lesions may contribute to alter the mutational landscape and purge cells accruing excess mutational load and/or DNA adducts that cannot be processed. In the case of BE4max, the transient overexpression of UGI and consequent inhibition of UG might impair the processing of spontaneous and induced cytidine deamination, preventing initiation of the endogenous BER and leading to engagement of the less faithful MMR or NHEJ, which may allow incorporation of transversions and trigger DNA damage response and apoptosis (Li, 1999; Sobol *et al*, 2002). A broad mutation pattern is naturally installed by cytidine deaminases during somatic hypermutation, when MMR may interfere with BER because of excess U-G mismatches (Schanz *et al*, 2009) . Of note, CBEs have been also previously reported to occasionally install transversions at the target site, with variable frequencies depending on the loci and the cell types (Kurt *et al*, 2021b; Zeng *et al*, 2020a), despite the underlying mechanisms remain unclear. In addition, the relatively high frequency of DNA DSBs induced at gRNA-dependent target sites of BE4max also causes p53 activation leading to loss of engrafting capacity and may thus purge the cells that have experienced highest exposure to the BE. Both processes may result in depletion of C>T/G>A transitions and provide an indirect readout of interference with normal DNA repair processes. Notably, when induction of DNA DSBs was alleviated by improved BE expression, the expected increase in these transitions appeared to emerge. In the case of ABE8.20-m, where no specific excision and repair pathway exists for DNA embedded inosines, one can expect engagement of MMR or NHEJ in absence of a concurrent DNA nick on the opposite strand, as it occurs instead at the target site. Error-prone repair may thus emerge as BE expression is increased, as noted in our experiments. As experimental conditions might alleviate or aggravate such impact, as shown here when treating cells at different culture times or using BE expression constructs with different efficiency, further

studies to investigate the mechanism(s) underlying any global impact of BEs and to devise strategies circumventing them are recommended for a comprehensive assessment of the risk benefit associated with these technologies.

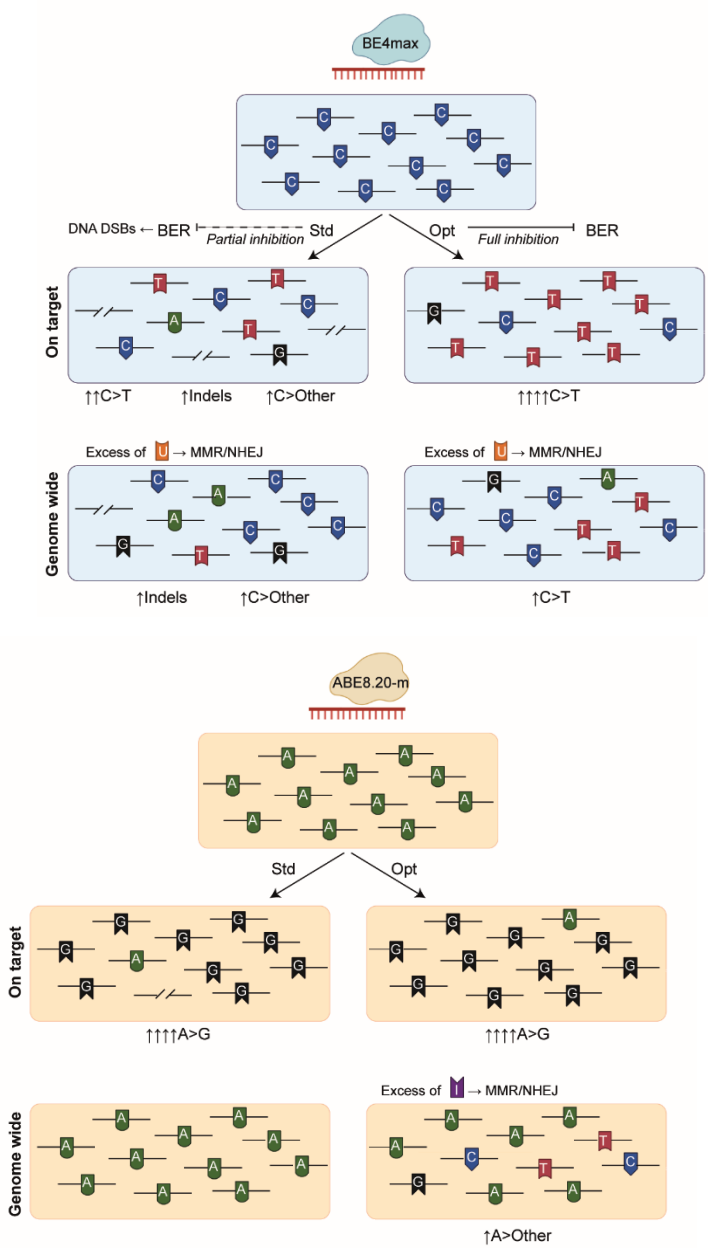


Figure 35. Schematic representation of the possible molecular mechanisms leading to mutagenic events at the target sites and genome wide upon base editing.

Like for BE, the constitutive albeit transient expression of the RT domain in case of prime editing raises concerns about its possible off-target activity. Although RT activity and its extent are challenging to be measured, this evaluation will be crucial to fully understand the genotoxic risk of PE. A possible approach to capture these convincible

rare and random events is to use ultra-high coverage whole genome sequencing techniques on a limited number of edited clones, such as those retrieved from xenograft or from pool of colonies.

6.4 Upsides and pitfalls of SMArT strategies in human HSPCs

In the second part of our study, our objective was to devise strategies for expanding the applications of HDR editing in human HSPCs. Several pathological conditions may necessitate different levels of corrected HSPCs to rescue the disease phenotype. This requirement can be contingent upon whether edited HSPCs gain a selective advantage or not. In general, even a modest proportion of corrected cells can yield clinical benefits if these cells acquire a selective advantage. For instance, 10% edited HSPCs in presence of myeloablative conditioning fully rescue SCID-X1 disease (Schiroli *et al*, 2017). Conversely, when corrected cells do not bear a selective advantage, the efficiency of editing required for therapeutic benefit cannot be sometimes reached due to the HDR constraints in human HSPCs. We thus envisioned to enrich a pure population of cells bearing the intended modification. The two selection strategies SMArT-2 and SMArT-3 validated in this project allow to select for engrafting HDR edited HSPCs and to purge out DNA DSB genotoxic byproducts occurring at the target site. Selected cells were capable of long-term engraftment upon xenotransplantation while still displaying high levels of on-target efficiency.

In the past years, Dever and colleagues devised a selection approach to enrich for *β-globin* HDR-edited HSPCs by the insertion of a constitutively expressed selector cassette in the *HBB* locus. Yet, the stable expression of the selector cassette raises concerns since it may induce differentiation, toxicity or immune response against the transgene (Dever *et al*, 2016). The primary difference between our SMArT-2 and SMArT-3 strategies and the previously published approach lies in the transient nature of our selector gene expression. In our strategies, the selector gene is highly active during the selection process but then shifts to a basal level once the mRNA encoding the transactivator is degraded. This design minimizes the potential complications associated with prolonged selector expression. However, in SMArT-2 strategy selector expression, which is poorly detectable at basal level in primitive HSPCs as the target gene is not expressed, can be expressed upon cell differentiation in some mature cells along with the edited gene. Similarly, in SMArT-3 strategy the use of strong promoter driving the expression of the

gene of interest may in principle transactivate the minimal promoter of the selector gene, even if we did not detect those effect when using the PGK promoter. One significant advantage of the SMArT strategies compared to previously published enrichment strategies (Mikkelsen *et al*, 2023) is the ability to enrich cells that have undergone on-target integration of the editing cassette. This advantage arises from the ArTs DBD binding to the genomic sequences flanking the site of on-target integration. Consequently, when the construct integrates into an off-target site, the selector gene remains inactive, leading to purge out of these cells. However, it is plausible that individual cells harboring both on-target and off-target integrations may still be positively selected. Finally, SMArT strategies allowed to purge out DNA DSB byproducts which represent the major concern when aiming to translate an editing approach into clinical practice. In selector positive cells, we noticed a noteworthy trend towards a decrease in large deletions when examining the region bound by ArTs. This trend becomes more evident when utilizing male cells and targeting an X-linked locus, as in the case of SMArT-2 approach. Indeed, when targeting a biallelic gene, the mixed outcomes from the two alleles impedes the complete depletion of cells bearing genotoxic byproducts as the inadvertently edited allele may often co-segregate with the HDR-edited one.

LT-HSCs exhibit a reduced propensity for repairing DSB by HDR than progenitors. Additionally, we uncovered that they express the selector gene to lower levels compared to the progenitor counterpart. These features bias the enrichment process toward progenitor cells at the expense of LT-HSCs, resulting in a lower long-term engraftment of positively enriched cells compared to the bulk population. Although a high number of injected progenitors could support the initial phase of engraftment and facilitate immune reconstitution immediately after transplantation, a lower number of injected LT-HSCs may raise concerns about long-term oligoclonal reconstitution in presence of myeloablative conditioning. To overcome this limitation, enrichment strategies can be combined with non-myeloablative (Omer-Javed *et al*, 2022a) or partial conditioning regimens as well as with advanced culture conditions that better preserve HSPCs during *in vitro* manipulation an allow expansion of primitive HSC (Rai *et al*, 2023; Bai *et al*, 2019a; Sakurai *et al*, 2023). Furthermore, our study demonstrates the feasibility of combining a selection-based approach with enhanced gene editing techniques. This combined approach increases the pool of corrected clones available for selection,

ultimately mitigating the challenges posed by oligoclonal reconstitution. The combination of all these approaches holds the promise of substantial benefits in SMArT strategies possibly leading to enhanced engraftment outcomes following transplantation.

In conclusion, the blueprint, set of metrics, and the innovative editing approaches outlined in this study will serve as valuable guidelines for the comprehensive assessment of editing systems. Additionally, they will expand the potential applications of gene editing in HSPCs, facilitating their transition towards clinical translation.

7 MATERIALS AND METHODS

7.1 Plasmids

The pCMV_BE4max plasmid was a gift from David Liu (Addgene plasmid: #112093; http://n2t.net/addgene:112093). The ABE8.20-m plasmid was a gift from Nicole Gaudelli (Addgene plasmid #136300; http://n2t.net/addgene:136300). The pCMV-PE2 plasmid was a gift from David Liu (Addgene plasmid #132775; http://n2t.net/addgene:132775). pCMV-PEmax was a gift from David Liu (Addgene plasmid # 174820; http://n2t.net/addgene:174820; RRID:Addgene_174820). The Cas9_WPRE-polyA and dCas9_WPRE-polyA plasmids were a gift from Angelo Lombardo (SR-Tiget). The nCas9, dCas9-RT and nCas9-dRT were obtained by mutagenesis from Cas9_WPRE-polyA and pCMV-PE2 plasmid respectively. All plasmids carried the T7 promoter downstream the CMV promoter. For the generation of constructs for standard (Std) mRNA in vitro transcription (IVT), the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE) followed by a polyA sequence were subcloned into the above-mentioned plasmids downstream the coding sequence. For the generation of constructs for optimized (Opt) mRNA IVT, the following sequences were subcloned in the Std plasmid in place of theT7 sequence for ARCA capping and the 5'UTR: CapAG – eIF4G aptamer (GACTCACTATTTGTTTTCGCGCCCAGTTGCAAAAAGTGTCG) - Kozak sequence (CCACC) – start codon (ATG), as previously described in(Omer-Javed *et al*, 2022b). The pU6-pegRNA-GG-acceptor was a gift from David Liu (Addgene plasmid #132777; http://n2t.net/addgene:132777). pegRNAs targeting *B2M* were designed with pegFinder (http://pegfinder.sidichenlab.org/) (Chow *et al*, 2020) using the

default parameters. pegRNAs were subcloned in the pU6-pegRNA-GG-acceptor using annealed oligonucleotides.

Description	Orientation	Sequence 5'-3'
peg1/2 spacer	Fw	caccgCAGCCCAAGATAGTTAAGTGgttttaga
	Rv	tagctctaaaacCACTTAACTATCTTGGGCTGc
peg1 RT + PBS	Fw	gtgcAAGACTTGCCCCACTTAACTATCTTGGGC
	Rv	aaaaGCCCAAGATAGTTAAGTGGGGCAAGTCTT
peg1/2 sgRNA	Fw	caccgAGTCACATGGTTCACACGGC
	Rv	aaacGCCGTGTGAACCATGTGACTc
peg2 RT + PBS	Fw	gtgcAAGACTTACCTCACTTAACTATCTTGGGC
	Rv	aaaaGCCCAAGATAGTTAAGTGAGGTAAGTCTT
peg3 spacer	Fw	caccgAGTCACATGGTTCACACGGCgttttaga
	Rv	tagctctaaaacGCCGTGTGAACCATGTGACTc
peg3 RT+ PBS	Fw	gtgcACTTAAAAAGATGAGTATGCCTGCCGTGTGAACCATGT
	Rv	aaaaACATGGTTCACACGGCAGGCATACTCATCTTTTTAAGT
peg3 sgRNA	Fw	caccgACTTGTCTTTCAGCAAGGAC
	Rv	aaacGTCCTTGCTGAAAGACAAGTc
peg4 spacer	Fw	caccgCTGAATCTTTGGAGTACCTGgttttaga
	Rv	tagctctaaaacCAGGTACTCCAAAGATTCAGc
peg4 RT + PBS	Fw	gtgcGATATTCCTCAAGTACTCCAAAGATTC
	Rv	aaaaGAATCTTTGGAGTACTTGAGGAATATC
peg4 sgRNA	Fw	caccgTCACGTCATCCAGCAGAGAA
	Rv	aaacTTCTCTGCTGGATGACGTGAc
peg5 spacer	Fw	caccGAGTAGCGCGAGCACAGCTAgttttaga
	Rv	tagctctaaaacTAGCTGTGCTCGCGCTACTC
peg5 RT+ PBS	Fw	gtgcTCCGTGGCCTGAGCTGTGCTCGCGC
	Rv	aaaaGCGCGAGCACAGCTCAGGCCACGGA
peg5 sgRNA	Fw	caccgAGTGGAGGCGTCGCGCTGGC
	Rv	aaacGCCAGCGCGACGCCTCCACTc
peg scaffold	Fw	5' Phos-GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCACCGAGTCG
	Rv	5' Phos-GCACCGACTCGGTGCCACTTTTTCAAGTTGATAA CGGACTAGCCTTATTTTAACTTGCTATTTC

For SMArT strategies the donor constructs were cloned using molecular biology technique, using different restriction enzymes (New England Biolabs Inc) and ligated with Quick LigationTM Kit (New England BioLabs Inc). Plasmid DNA (0,5-5ng for closed plasmids or 30-100ng for ligation products) to be transformed was mixed with 50-100µl of chemically competent Top10 cells (Invitrogen). The mix was incubated on ice for 30 minutes, then 45 seconds in a water bath at 42°C, and then returned to ice for a further 2 minutes. Transformed bacteria were grown for 30 min in LB (for ligation products) or directly (for closed plasmids) plated onto Luria Bertani (LB) agar plates, containing 100µg/ml ampicillin or kanamycin according to antibiotic resistance

expression of the plasmid, and placed overnight in a 37°C bacterial incubator. The following day, colonies were picked and placed into 3 ml of LB media containing carbenicillin (ampicillin homologue) or kanamycin and grown over day with agitation at 37°C. DNA was extracted using Wizard® Plus SV Minipreps DNA Purification System (Promega) and enzymatic digestion was performed to screen for positive clones. 200µl of the positive culture were placed in 500ml of LB media containing carbenicillin or kanamycin and grown overnight shaking at 37°C. The following day, cell suspension was measured by spectrophotometer analysis (λ=600nm) to check for optimal cell density (about 1-1,5 O.D.) and subjected to large-scale plasmid DNA preparation with EndoFree Plasmid Maxi Kit (QIAGEN).

7.2 gRNAs, pegRNA and DNA binding site sequences

Sequences of the gRNAs were designed using an online CRISPR design tool and selected for predicted specificity score and on target activity. Genomic sequences recognized by the gRNAs are indicated below.

gRNA ID	gRNA Sequence 5'-3'
gRNA B2M ex2	CTTACCCCACTTAACTATCT
gRNA CD40LG (unrelated sgRNA)	TGGATGATTGCACTTTATCA
gRNA AAVS1 #1	GTCACCAATCCTGTCCCTAG
gRNA AAVS1 #2	GATAAGGAATCTGCCTAAC
gRNA B2M ex1	GAGTAGCGCGAGCACAGCTA
gRNA BCL11A	TTTATCACAGGCTCCAGGAA
gRNA CCR5 #1	GGTGTCGAAATGAGAAGAAG
gRNA CCR5 #2	TGACATCAATTATTATACAT
gRNA IL2RG	ACTGGCCATTACAATCATGT
pegRNA 5 B2M ex1	mG*mA*mG*UAGCGCGAGCAC AGCUAGUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGG CUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGU GCUCCGUGGCCUGAGCUGU GCUCGCGCUmU*mU*mU
nicking gRNA 5 B2M ex1	AGTGGAGGCGTCGCGCTGGC

TALE has been cloned by Golden Gate strategy in pUC plasmid and VPR domain from plasmid #63789 fused at the C-terminus.

TALE binding sites for *IL2RG* are listed below.

TALE ID	Target sequence
TALE #1	CCCAAAACAGTAGAGCT
TALE #2	TCCTTCCCAGGATCTAGGT
TALE #3	GTCACACAGCACATATTT

TALE binding sites for *AAVS1* are listed below.

TALE ID	Target sequence
TALE #5	TCCACCATCTCATGCCCCT
TALE #4	TCCACACGGACACCCCCCT
TALE #9	TCCTCCGTGCGTCAGTTTT
TALE #8	TGGTCCTGAGTTCTAACTT
TALE #3	TCCAAACTGCTTCTCCTCT
TALE #2	TCCCACCCCCTGCCAAGCT
TALE #7	TCCTCTCTGGCTCCATCGT
TALE #1	TCCATCGTAAGCAAACCTT
TALE #6	TCTAAGGTTTGCTTACGAT

7.3 AAV6 donor templates

AAV6 donor templates for HDR were generated from a construct containing AAV2 inverted terminal repeats, produced by triple-transfection method and purified by ultracentrifugation on a cesium chloride gradient as previously described.

7.4 mRNA in vitro transcription

Std and Opt plasmids were linearized with SpeI or PmeI (New England Biolabs) and purified by phenol-chloroform extraction. Different preps of mRNAs for each editor were in vitro transcribed using the commercial 5X MEGAscript T7 kit (Thermo Fisher). Std mRNAs were capped with 4.5 mM Anti-Reverse Cap Analog (ARCA) 3′-O-Me-mG(5′) ppp(5′)G (New England Biolabs) mixed in a 1:5 ratio with dGTP nucleotides. Opt mRNAs were capped with 8mM CleanCapAG (Trilink) (Omer-Javed *et al*, 2022a). mRNAs were purified using RNeasy Plus Mini Kit (Qiagen). mRNAs were denaturated and resolved by capillary electrophoresis on 4200 TapeStation System (Agilent) according to the manufacturer's instructions to assess quality and integrity. mRNAs were purified by high-performance liquid chromatography (ADS BIOTEC WAVE System) and concentrated with Amicon Ultra-15 (30 K) tubes (Millipore). mRNA productions were then aliquoted and stored at −80 °C. Reproducible results were obtained in replicate experiments using different preparations of the same editor mRNA.

7.5 Cell lines and primary cell culture

B-lymphoblastoid cells were cultured in RPMI 1640 medium (Corning) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Euroclone), 100 IU ml−1 penicillin, 100 µg ml−1 streptomycin and 2% glutamine.

K-562 cells (ATCC) were cultured in Iscove's modified Dulbecco's medium (Corning) supplemented with 10% heat-inactivated FBS, 100 IU ml−1 penicillin, 100 µg ml−1 streptomycin and 2% glutamine.

Human primary T cells were isolated from healthy donors' peripheral blood mononuclear cells (PBMCs) freshly purified from buffy coats with SepMate™ PBMC Isolation Tubes (StemCell Technologies), according to manufacturing instructions. Buffy coats were obtained in accordance with the Declaration of Helsinki, as anonymized residues of blood donations, used upon signature of specific institutional informed consent for blood product donation by healthy blood donors. CD3+ T cells were stimulated using magnetic beads (1:3 cells to beads ratio) conjugated with anti-CD3/anti-CD28 antibodies (Dynabeads human T-activator CD3/CD28; Thermo Fisher). Cells were maintained in Iscove's modified Dulbecco's medium (Corning) supplemented with 10% heat-inactivated FBS, 100 IU ml−1 penicillin, 100 µg ml−1 streptomycin, 2% glutamine, 5 ng ml−1 human (h)IL-7 (PreproTech) and 5 ng ml−1 hIL-15 (PreproTech). Dynabeads were removed after 6 days of culture.

CB CD34+ HSPCs were purchased from Lonza according to the TIGET-HPCT protocol approved by OSR Ethical Committee and seeded at the concentration of 5x105 cells ml−1 in serum-free StemSpan SFEM (STEMCELL Technologies) supplemented with 100 IU ml−1 penicillin, 100 µg ml−1 streptomycin, 2% glutamine, 100 ng ml−1 hSCF (PeproTech), 100 ng ml−1 hFlt3-L (PeproTech), 20 ng ml−1 hTPO (PeproTech) and 20 ng ml−1 hIL-6 (PeproTech), 10 µM 16,16-Dimethyl Prostaglandin E2 (added at the beginning of the culture; Cayman), 1 µM SR1 (Biovision) and 50nM UM171 (STEMCELL Technologies).

G-CSF or G-CSF + Plerixafor mPB CD34+ HSPCs were purified in house with the CliniMACS CD34 Reagent System (Miltenyi Biotec) from Mobilized Leukopak (AllCells) according to the TIGET-HPCT protocol approved by OSR Ethical Committee and following the manufacturer's instructions. HSPCs were seeded at the concentration

of 5x105 cells ml−1 in serum-free StemSpan SFEM supplemented with 100 IU ml−1 penicillin, 100 µg ml−1 streptomycin, 2% glutamine, 300 ng ml−1 hSCF, 300 ng ml−1 hFlt3-L, 100 ng ml−1 hTPO and 10 µM 16,16-Dimethyl Prostaglandin E2 (added at the beginning of the culture), 1 µM SR1 and 35 nM UM171.

All cells were cultured in a 5% CO2 humidified atmosphere at 37 °C.

7.6 Gene editing of cell lines and analyses

For each condition, 3.0x105 cells were washed with ten volumes of Dulbecco's phosphate-buffered saline without Ca2+ and Mg2+ (DPBS; Corning) and electroporated using the SF Cell Line 4D-Nucleofector™ X Kit (Lonza). B-lymphoblastoid cells were pulsed with program EW-113, K-562 cells were pulsed with program FF-120. For base editing, B-lymphoblastoid cells were electroporated with 0.5 µg B2M gRNA plasmid and either 2.0 µg BE plasmid (Addgene) or 4.0 µg Std BE mRNAs unless otherwise specified. For PE, K-562 were electroporated with 0.25 µg pegRNA plasmids and 1 µg PE2 plasmid. For PE3 conditions, 0.25 µg of respective gRNA plasmids were added to the electroporation mixture. Cells were cultured 7 days, analyzed by flow cytometry and collected for genomic (gDNA) extraction and subsequent molecular analysis.

7.7 Gene editing of human T cells and analyses

For each condition, 5.0x105-1.0x106 human T cells were washed with ten volumes of DPBS without Ca2+ and Mg2+ and electroporated using the P3 Primary Cell 4D-Nucleofector X Kit (Lonza) and program DS-130. Cells were electroporated with 75 pM B2M gRNA (Synthego) and 3.0 µg Std mRNAs unless otherwise specified. Cells were cultured 7 days and then analyzed by flow cytometry and collected for genomic (gDNA) extraction and subsequent molecular analysis.

7.8 Gene Editing of Human CD34+ cells

For each condition, from 2.0x105 to 7.5x105 CB-/mPB-derived HSPCs were washed with ten volumes of DPBS without Ca2+ and Mg2+ and electroporated using the P3 Primary Cell 4D-Nucleofector X Kit (Lonza) and program EO-100 after either 1 or 3 days of culture, as indicated. For base and Cas9 editing, HSPCs were electroporated with 75 pmol gRNA (B2M ex2, AAVS1, B2M ex1, BLC11a, CCR5, IL2RG) and 7.5 µg Std mRNAs, unless otherwise specified, or 3.5 µg (LD) or 7.5 µg (HD) Opt mRNAs. For PE, HSPCs were electroporated with 186 pmol B2M pegRNA5, 75 pmol B2M exon 1 gRNA

and 7.5 μg Opt PE mRNA. Seven days after electroporation HSPCs were collected for flow cytometry analyses and to extract genomic gDNA for molecular analysis. Colony-forming cell (CFC) assays were performed 24 hrs after editing procedure by plating 400-800 cells in methylcellulose-based medium (MethoCult H4434, StemCell Technologies) supplemented with 100 IU ml−1 penicillin and 100 μg ml−1 streptomycin. Three technical replicates were performed for each condition. Two weeks after plating, colonies were counted and eventually picked for molecular analysis or exome sequencing.

7.9 Mice

All experiments and procedures involving animals were performed with the approval of the Animal Care and Use Committee of the San Raffaele Hospital (IACUC no. 1206) and authorized by the Italian Ministry of Health and local authorities accordingly to Italian law. NOD-SCID-IL2Rg−/− (NSG) female mice (The Jackson Laboratory) were held in specific pathogen-free conditions.

7.10 CD34+ HSPC xenotransplantation experiments in NSG mice

In base and prime editing experiments for xenotransplantation of CB and G-CSF mPB HSPCs, the outgrowths of 5.0x104 to 1.25x105 and 5.0x105 to 7.5x105 HSPCs, respectively, at the start of the culture (t0 equivalent) were injected intravenously 24 hrs after editing into sublethally irradiated NSG mice (180-200 cGy). Matched numbers of HSPCs were seeded at day 0 of culture for each experimental group in order to transplant in each mouse the same number of culture-initiating HSPCs. Mice were randomly distributed to each experimental group. Human CD45+ cell engraftment and the presence of edited cells were monitored by serial collection of blood (approximately every 2 to 3 weeks) from the retroorbital plexus and, at the end of the experiment (15-16 weeks after transplantation), BM and SPL were collected for end-point analyses, including florescence-activated cell sorting (FACS) of hematopoietic lineages in some experiments. Secondary transplantation in NSG mice was performed by transplanting beads purified (Miltenyi Biotec) human CD34+ cells from BM of primary recipients. CD34+ cells from all mice of each experimental group were pooled and split in recipients, according to the input number of cells.

7.11 Flow cytometry

Immunophenotypic analyses were performed by flow cytometry using Canto II (BD Pharmingen). From 5.0x104 to 2.0x105 cells either from culture or mouse-derived samples were analyzed. Cells were stained for 15' at 4 °C with antibodies listed below in a final volume of 100 µl and then washed with DPBS + 2% heat-inactivated FBS. Single stained and fluorescence-minus-one-stained cells were used as controls. The Live/Dead Fixable Dead Cell Stain Kit (Thermo Fisher) or 7-aminoactinomycin D (Sigma Aldrich) were included during sample preparation according to the manufacturer's instructions to identify dead cells. Apoptosis analysis was performed on T cells one day after electroporation using Pacific Blue-conjugated Annexin V (Biolegend) and the Apoptosis Detection kit with 7-Aminoactinomycin D (7AAD, BD Pharmingen) according to the manufacturers' instructions. Percentages of live (7AAD−, AnnexinV−), early apoptotic (7AAD−, AnnexinV+), late apoptotic (7AAD+, AnnexinV+) and necrotic (7AAD+, AnnexinV−) cells were reported. Cell sorting was performed on a BD FACSAria Fusion (BD Biosciences) using BD FACS Diva software v8.0.1 and equipped with four lasers: blue (488 nm), yellow/green (561 nm), red (640 nm) and violet (405 nm). Cells were sorted with an 85 µm nozzle. Sheath fluid pressure was set at 45 psi. A highly pure sorting modality (four-way purity sorting) was chosen. Cell sorting was performed on a MoFlo Astrios EQ (Beckman Coulter) using Summit software and equipped with four lasers: blue (488 nm), yellow/green (561 nm), red (640 nm) and violet (405 nm). Cells were sorted with a 100 µm nozzle. Sheath fluid pressure was set at 25 psi. A highly pure sorting modality (purify-1 sorting) was chosen. Sorted cells were collected in 1.5 ml Eppendorf tubes containing 500 µl of DPBS. Data were analyzed with FCS Express 7 Flow.

Antibody	Conjugate	Clone	Supplier	Identifier	Dilution
Anti-human CD34	VioBlue	AC136	Miltenyi Biotec	Catalog n°130-113-182	1:50
Anti-human CD133/2	PE	REA816	Miltenyi Biotec	Catalog n°130-112-157	1:50
Anti-human CD90	APC	5E10	BD Biosciences	Catalog n°559869	1:33
Anti-human B2M	Pecy7	2M2	Biolegend	Catalog n°316318	1:100
Anti-human CD45	VioBlue	HI30	BioLegend	Catalog n°304029	1:50
Anti-human CD45	APC	2D1	BD Biosciences	Catalog n°340910	1:50
Anti-human CD45	APC-Vio770	2D1	BD Biosciences	Catalog n°348815	1:50
Anti-human CD19	PE	SJ25C1	BD Biosciences	Catalog n°345789	1:50
Anti-human CD3	APC	UCHT1	BD Biosciences	Catalog n° 555335	1:50
Anti-human CD3	FITC	SK7	BD Biosciences	Catalog n° 345763	1:50
Anti-human CD13	APC	WM15	BD Biosciences	Catalog n°557454	1:50
Anti-human CD13	BV421	WM15	BD Biosciences	Catalog n°562596	1:50
Anti-human CD33	PE-Vio770	WM15	Miltenyi Biotec	Catalog n°130-113-350	1:50
Anti-human CD38	PerCP/Cyanine5.5	HB-7	BioLegend	Catalog n°356614	1:20
Anti-human CD271(NGFR)	PE	ME20.4-1.H4	Miltenyi Biotec	Catalog n°130-113-421	1:50
Anti-human CD271(NGFR)	APC	ME20.4-1.H4	Miltenyi Biotec	Catalog n°130-113-418	1:50
human Fc blocking			Miltenyi Biotec	Catalog n°130-092-283	1:50
mouse Fc blocking			BD Biosciences	Catalog n°553142	1:100

7.12 Molecular analyses

For molecular analyses, gDNA was isolated with QIAamp DNA Micro Kit (QIAGEN) according to the manufacturer's instructions. Extraction of genomic DNA from colonies in CFC assays was performed with QuickExtract (Epicentre) according to the manufacturer's instructions. When specified, BE and Cas9 efficiencies were measured by PCR amplification at the target locus followed by amplicon Sanger sequencing (Eurofins Scientific), whose results were then analyzed by EditR software (http://baseeditr.com) (Kluesner *et al*, 2018) using default parameters or by TIDE software (https://tide.nki.nl/) (Brinkman *et al*, 2014). When specified, PE efficiencies were measured by PCR amplification at the target locus followed by amplicon Sanger sequencing (Eurofins Scientific), whose results were then analyzed by EditR software. To adapt EditR for B2M prime editing, we used as input the sequence TGGCCTTAGCTGTGCTCGC selecting the reverse complement orientation option.

Description	Orientation	Sequence 5'-3'
B2M BE (ex2)	Fw	GTCATCCAGCAGAGAATGGAAA
	Rv	AGTAGGTAAGAAGTGTTAAGAGTGT
AAVS1#1 BE	Fw	CTTCAGGACAGCATGTTTGC
	Rv	GGACTAGAAAGGTGAAGAGCC
B2M PE (ex1)	Fw	CGCGTTTAATATAAGTGGAGGC
	Rv	GGAGAACTTGGAGAAGGGAAGT

For digital droplet PCR (ddPCR) analyses, 5–50 ng of gDNA was analyzed using the QX200 Droplet Digital PCR System (Bio-Rad) according to the manufacturer's instructions. Primers and probes for vector copy number (VCN) were previously reported (Soldi *et al*, 2020). Primers and probes to detect large B2M deletions were designed upstream and downstream the DNA SSB of base and prime editing or of Cas9 DSB. Human TTC5 (Bio-Rad) or GAPDH (Bio-Rad) assays were used for normalization. Copy numbers for both VCN and deletion analyses were calculated with the following formula: no. of LV/B2M+ droplets / no. of normalizer+ droplets multiplied by 2.

Description	Orientation	Sequence 5'-3'
deletions BE ex2 5'	Fw	CCTGAATTGCTATGTGTC
	Rv	CAGTGTAGTACAAGAGATAGA
	Probe	CATCCATCCGACATTGAAGTTGACT
deletions BE ex2 3'	Fw	CCTGAATGAGTCCCATCCCA
	Rv	ACTGCAGGGAAACTACTGGT
	Probe	TGGCACCTGCTGAGATACTGATGCA
deletions BE ex1 5'	Fw	GCCGATGTACAGACAGCAAA
	Rv	GCAGTGCCAGGTTAGAGAGA
	Probe	CTCACCCAGTCTAGTGCATGCCTTCT
deletions BE ex1 3'	Fw	CCGTGACTTCCCTTCTCCAA
	Rv	GATCCAGCCCTGGACTAGC
	Probe	TCTCCTTGGTGGCCCGCCGT
deletions PE 5'	Fw	same as deletions BEex1 5'
	Rv	
	Probe	
deletions PE 3'	Fw	same as deletions BEex1 3'
	Rv	
	Probe	
IL2RG deletions	Fw	AGGTGGTTGAGAATGGTGCT
	Rv	CTGTGTGACACGGGCTAAGT
	Probe	CCTCAGCCCACCTAGATCCTGGG
AAVS1 deletions	Fw	ATCCTGGGAGGGAGAGCTT
	Rv	GGATCAGTGAAACGCACCAG
	Probe	AGCTGCTCTGACGCGGCCGT

For translocation analyses 100ng of gDNA were amplified. DNA amplicons were resolved by capillary electrophoresis on 4200 TapeStation (Agilent) according to the manufacturer's instructions. For gene expression analyses, total RNA was extracted using RNeasy Plus Micro Kit (QIAGEN), according to the manufacturer's instructions. DNAse

treatment was performed using RNase-free DNAse Set (QIAGEN). Complementary (c)DNA was synthesized using SuperScript VILO IV cDNA Synthesis Kit (Thermo Fisher) with EzDNAse treatment. 2 ng of cDNA were then used for gene expression by ddPCR. Relative expression of each target gene was first normalized to HPRT and then represented as fold changes relative to UT cells.

HPRT1	Probe (VIC)	Hs02800695_m1 (Life Technologies)
IRF7	Probe (FAM)	Hs01014809_g1 (Life Technologies)
OAS1	Probe (FAM)	Hs00973637_m1 (Life Technologies)
DDX58	Probe (FAM)	Hs00204833_m1 (Life Technologies)
CDKN1A	Probe (FAM)	Hs00355782_m1 (Life Technologies)
APOBEC3H	Probe (FAM)	Hs00419665_m1 (Life Technologies)
TP73	Probe (FAM)	Hs01056231_m1 (Life Technologies)
MT2A	Probe (FAM)	Hs01591333_g1 (Life Technologies)

For HDR ddPCR, primers and probes were designed on the junction between the vector sequence and the targeted locus and on control sequences used for normalization.

Description	**Orientation**	**Sequence 5'-3'**
Intron 1 IL2RG 3' integration junction ddPCR	FW	CTAGATTGGGGAGAAAATGA
	RV	GTGGGAAGGGGCCGTACAG
	Probe (FAM)	GTAGCTCCTATGCTAGGCGTAGCC
AAVS1 3' integration junction ddPCR	FW	GATTGGGAAGACAATAGCAG
	RV	TCTTGGGAAGTGTAAGGAAG
	Probe (FAM)	CCAGATAAGGAATCTGCCTA

Digital droplet PCR data were analyzed with QuantaSoftTM Software v1.7.4 (Bio-Rad).

7.13 Deep sequencing and bioinformatic analyses

PCR amplicons for individual samples were generated by nested PCR starting from >50-100 ng of purified gDNA. List of primers and thermal cycling protocol are reported in (Fiumara *et al*, 2023). For B2M ex2, AAVS1, B2M ex1, BCL11A and IL2RG the first PCR step was performed with GoTaq G2 DNA Polymerase (Promega) according to manufacturer instruction. The second PCR step was performed using the same reagents of the first step and5 µl of the PCR, For CCR5 base editing, a pre-amplification step followed by first and second PCR were performed with GoTaq G2 DNA Polymerase (Promega) according to manufacturer instruction. Primers used for the second PCR step

contained P5/P7 sequences, i5/i7 Illumina tags to allow multiplexed sequencing and R1/R2 primer binding sites. The PCR amplicon from each sample was separately purified by QIAquick PCR Purification Kit (QIAGEN). Concentration and quality of amplicons were assessed by QuantiFluor ONE dsDNA system and 4200 Tapestation System (Agilent). Amplicons from up to 49 differently tagged samples were multiplexed at equimolar ratios and run by the San Raffaele Center for Omic Sciences (COSR) using 1 x 150 bp paired-end MiSeq (Illumina). Sequencing data were analyzed with CRISPResso2 (v2.2.8), which enables the detection of small variants in gene editing experiments (Clement *et al*, 2019)

7.14 HSPCs Clonal Tracking and analysis

The transfer vector construct for the barcoded LV (BAR LV) was kindly provided by Dr. Bernhard Gentner. The LV was produced as described in (Soldi *et al*, 2020). Clonal tracking was performed on a pool of HSPCs derived from four CB donors. One day after thawing, HSPCs were transduced at the concentration of 1x10^6 cells ml−1 with the BAR LV using a multiplicity of infection of 30 transducing units ml−1. HSPCs were washed with ten volumes of DPBS without Ca2+ and Mg2+ 24 hrs later and then (24 hrs after washing) treated for editing (or mock electroporated) and transplanted as described above. 16 weeks after transplantation hematopoietic lineages were sorted from hematopoietic organs. gDNA was extracted and prepared for deep sequencing as described above. Sequencing data were analyzed with the BAR-Seq2 pipeline (https://bitbucket.org/bereste/bar-seq2).

7.15 Total RNA-Seq library preparation and analysis

Whole transcriptomic analysis was performed on a pool of HSPCs derived from six CB donors. All conditions were performed in triplicate. Total RNA was purified 24 hrs after editing using RNeasy Micro Kit (QIAGEN). DNase treatment was performed using RNase-free DNAse Set (QIAGEN), according to the manufacturer's instructions. RNA was quantified with Qubit 2.0 Fluorometer (Thermo Fisher) and its quality assessed by a 2100 Agilent Bioanalyser (Agilent Technologies). Minimum quality was defined as RNA integrity number (RIN) > 8. 300 ng of total RNA were used for library preparation with TruSeq Stranded mRNA (Illumina) and sequenced on a NextSeq 500 High 75 (Illumina) by the San Raffaele Center for Omic Sciences (COSR) or Genewiz (Azenta Life

Sciences). Pre-processing of the input sequences was done with FastQC (v0.11.6) to assess reads quality and trimmomatic to get rid of low-quality sequences. Then, reads were aligned to the human genome assembly (GRCh38) using the STAR software (v2.7.6a) with standard parameters, and abundancies were calculated using the Subread featureCounts function (v2.0.1). Differential Gene Expression (DGE) analysis was performed using the R/Bioconductor package DESeq2 (v1.30.0), normalizing for library size using DESeq2's median of ratios. p-values were corrected using false discovery rate (FDR), and genes having FDR < 0.05 were considered as differentially expressed. Post-analyses on DGE results were performed with the R/Bioconductor package ClusterProfiler (v4.7.1), using the Hallmark collection from MSigDB as reference database. Visualization of the (spliced) alignments on the TP73 gene was done with Integrative Genomes Viewer (IGV v2.8.0).

7.16 Whole Exome Sequencing (WES) for the detection of gRNA-independent DNA off-targets

For WES in Fig. 21, CB-derived HSPCs were edited as described above for the clonal tracking experiment and collected 7 days after the procedure to perform 500X WES of the in vitro bulk population. Cells from the same treatments were infused one day after treatment in NSG mice and live human CD45+ cells from BM were retrieved 16 weeks after infusion for 500X WES. For mock electro mice, live human CD45+ cells were sorted, gDNA was extracted and sequencing was performed as described below. For WES in Fig. 25a-e, mPB-derived HSPCs were edited and collected 7 days after the procedure to perform 500X WES of the in vitro bulk population. Cells from the same treatments were infused one day after treatment in NSG mice and live human CD45+ cells from BM were retrieved 16 weeks after infusion for 500X WES. For mock electro mice, live human CD45+ cells were sorted, gDNA was extracted and sequencing was performed as described below. For WES in Fig. 25f-i, mPB-derived HSPCs from one donor were treated and plated 24 hrs later for CFC assay. The bulk mock electro sample was also collected 24 hrs after editing and sequenced by 100X WES. Two weeks later individual colonies were picked, screened for intended outcome, and six single colonies for each condition were pooled in equal gDNA amount and sequenced by WES 100X. All WES were performed by Genewiz (Azenta) using the Agilent SureSelect Human All Exon V7 kit and running an Illumina NovaSeq 2x150bp with an estimated output of ~50Gb (500X)

or ~10Gb (100X) per sample. WES data were analyzed following the GATK "Best Practice Workflows" to identify variants in each sample.

7.17 Statistical analyses

The number of biologically independent samples, animals or experiments is indicated by “n”. For some experiments, different HSPC donors were pooled to account for donor-related variability and reach the number of cells needed for the analyses. Data were summarized as median with IQR (or range) or mean ± s.e.m. depending on data distribution. Inferential techniques were applied in presence of adequate sample sizes ($n \geq 5$), otherwise only descriptive statistics are reported.

8 REFERENCES

Abarrategi A, Foster K, Hamilton A, Mian SA, Passaro D, Gribben J, Mufti G & Bonnet D (2017) Versatile humanized niche model enables study of normal and malignant human hematopoiesis. *J Clin Invest* 127: 543–548

Adigbli G, Hua P, Uchiyama M, Roberts I, Hester J, Watt SM & Issa F (2021) Development of LT-HSC-Reconstituted Non-Irradiated NBSGW Mice for the Study of Human Hematopoiesis In Vivo. *Front Immunol* 12: 642198

Adikusuma F, Piltz S, Corbett MA, Turvey M, McColl SR, Helbig KJ, Beard MR, Hughes J, Pomerantz RT & Thomas PQ (2018) Large deletions induced by Cas9 cleavage. *Nature 2018 560:7717* 560: E8–E9

Adli M (2018) The CRISPR tool kit for genome editing and beyond. *Nat Commun* doi:10.1038/s41467-018-04252-2 [PREPRINT]

Al-Bahlani S, Fraser M, Wong AYC, Sayan BS, Bergeron R, Melino G & Tsang BK (2011) P73 regulates cisplatin-induced apoptosis in ovarian cancer cells via a calcium/calpain-dependent mechanism. *Oncogene* 30: 4219–4230

Allen D, Kalter N, Rosenberg M & Hendel A (2023) Homology-Directed-Repair-Based Genome Editing in HSPCs for the Treatment of Inborn Errors of Immunity and Blood Disorders. *Pharmaceutics* 15

Altuntaş F & Korkmaz S (2017) Hematopoietic progenitor cell mobilization. *Transfusion and Apheresis Science*

Amabile A, Migliara A, Capasso P, Biffi M, Cittaro D, Naldini L & Lombardo A (2016) Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell*

Antoniou P, Hardouin G, Martinucci P, Frati G, Felix T, Chalumeau A, Fontana L, Martin J, Masson C, Brusson M, *et al* (2022) Base-editing-mediated dissection of a γ-globin cis-regulatory element for the therapeutic reactivation of fetal hemoglobin expression. *Nature Communications 2022 13:1* 13: 1–22

Anzalone A V., Gao XD, Podracky CJ, Nelson AT, Koblan LW, Raguram A, Levy JM, Mercer JAM & Liu DR (2022) Programmable deletion, replacement, integration and

inversion of large DNA sequences with twin prime editing. *Nat Biotechnol* 40: 731–740

Anzalone A V., Koblan LW & Liu DR (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology 2020 38:7* 38: 824–844

Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, *et al* (2019a) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*

Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, *et al* (2019b) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576: 149–157

Bai T, Li J, Sinclair A, Imren S, Merriam F, Sun F, O'Kelly MB, Nourigat C, Jain P, Delrow JJ, *et al* (2019a) Expansion of primitive human hematopoietic stem cells by culture in a zwitterionic hydrogel. *Nat Med*

Bai T, Li J, Sinclair A, Imren S, Merriam F, Sun F, O'Kelly MB, Nourigat C, Jain P, Delrow JJ, *et al* (2019b) Expansion of primitive human hematopoietic stem cells by culture in a zwitterionic hydrogel. *Nature Medicine 2019 25:10* 25: 1566–1575

Bai T, Li J, Sinclair A, Imren S, Merriam F, Sun F, O'Kelly MB, Nourigat C, Jain P, Delrow JJ, *et al* (2019c) Expansion of primitive human hematopoietic stem cells by culture in a zwitterionic hydrogel. *Nature Medicine 2019 25:10* 25: 1566–1575

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA & Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709–1712

Batista Napotnik T, Polajžer T & Miklavčič D (2021) Cell death due to electroporation - A review. *Bioelectrochemistry* 141

Baum CM, Weissman IL, Tsukamoto AS, Buckle AM & Peault B (1992) Isolation of a candidate human hematopoietic stem-cell population. *Proc Natl Acad Sci U S A*

Becker S & Boch J (2021) TALE and TALEN genome editing technologies. *Gene and Genome Editing* 2: 100007

Beerman I, Seita J, Inlay MA, Weissman IL & Rossi DJ (2014) Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* 15: 37

Benveniste P, Frelin C, Janmohamed S, Barbara M, Herrington R, Hyam D & Iscove NN (2010) Intermediate-Term Hematopoietic Stem Cells with Extended but Time-Limited Reconstitution Potential. *Cell Stem Cell*

Berenson RJ, Andrews RG, Bensinger WI, Kalamasz D, Knitter G, Buckner CD & Bernstein ID (1988) Antigen CD34+ marrow cells engraft lethally irradiated baboons. *Journal of Clinical Investigation*

Bhatia M, Wang JCY, Kapp U, Bonnet D & Dick JE (1997) Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci U S A*

Bizard AH & Hickson ID (2014) The Dissolution of Double Holliday Junctions. *Cold Spring Harb Perspect Biol* 6

Blackford AN & Jackson SP (2017) ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell* 66: 801–817

Boch J & Bonas U (2010) Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function . *Annu Rev Phytopathol*

Bogdanove AJ & Voytas DF (2011) TAL effectors: Customizable proteins for DNA targeting. *Science (1979)* doi:10.1126/science.1204094 [PREPRINT]

Boitano AE, Wang J, Romeo R, Bouchez LC, Parker AE, Sutton SE, Walker JR, Flaveny CA, Perdew GH, Denison MS, *et al* (2010) Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science (1979)*

Boulais PE & Frenette PS (2015) Making sense of hematopoietic stem cell niches. *Blood*

Braun CJ, Boztug K, Paruzynski A, Witzel M, Schwarzer A, Rothe M, Modlich U, Beier R, Göhring G, Steinemann D, *et al* (2014) Gene therapy for Wiskott-Aldrich syndrome-long-term efficacy and genotoxicity. *Sci Transl Med*

Brinkman EK, Chen T, Amendola M & Van Steensel B (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 42: e168–e168

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin E V. & Van Der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321: 960–964

Brown BD, Gentner B, Cantore A, Colleoni S, Amendola M, Zingale A, Baccarini A, Lazzari G, Galli C & Naldini L (2007) Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat Biotechnol*

Bunting SF, Callén E, Wong N, Chen HT, Polato F, Gunn A, Bothmer A, Feldhahn N, Fernandez-Capetillo O, Cao L, *et al* (2010) 53BP1 inhibits homologous recombination in brca1-deficient cells by blocking resection of DNA breaks. *Cell*

Burma S, Chen BPC & Chen DJ (2006) Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)* doi:10.1016/j.dnarep.2006.05.026 [PREPRINT]

Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S & Hoffmann C (2005) Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol* doi:10.1038/nrmicro1263 [PREPRINT]

Canarutto D, Tucci F, Gattillo S, Zambelli M, Calbi V, Gentner B, Ferrua F, Marktel S, Migliavacca M, Barzaghi F, *et al* (2021) Peripheral blood stem and progenitor cell collection in pediatric candidates for ex vivo gene therapy: a 10-year series. *Mol Ther Methods Clin Dev* 22: 76

Carrelha J, Meng Y, Kettyle LM, Luis TC, Norfo R, Alcolea V, Boukarabila H, Grasso F, Gambardella A, Grover A, *et al* (2018) Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* 554: 106–111

Cavazza A, Moiani A & Mavilio F (2013) Mechanisms of retroviral integration and mutagenesis. *Hum Gene Ther* doi:10.1089/hum.2012.203 [PREPRINT]

Chad M, Rivella S, Callegari J, Heller G, Gaensler KML, Luzzatto L & Sadelain M (2000) Therapeutic haemoglobin synthesis in β-thalassaemic mice expressing lentivirus-encoded human β-globin. *Nature*

Challita PM & Kohn DB (2006) Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. *Proceedings of the National Academy of Sciences*

Chang HHY, Pannunzio NR, Adachi N & Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology 2017 18:8* 18: 495–506

Charpentier M, Khedher AHY, Menoret S, Brion A, Lamribet K, Dardillac E, Boix C, Perrouault L, Tesson L, Geny S, *et al* (2018) CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat Commun*

Chatterjee N & Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 58: 235–263

Chen L, Park JE, Paa P, Rajakumar PD, Prekop HT, Chew YT, Manivannan SN & Chew WL (2021a) Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat Commun* 12

Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, Chen C, Nelson JW, Newby GA, Sahin M, *et al* (2021b) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* 184: 5635-5652.e29

Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, Chen C, Nelson JW, Newby GA, Sahin M, *et al* (2021c) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* 184: 5635-5652.e29

Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, Chen C, Nelson JW, Newby GA, Sahin M, *et al* (2021d) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* 184: 5635-5652.e29

Chen Z, Yang H & Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453: 489–494

Chow RD, Chen JS, Shen J & Chen S (2020) A web tool for the design of prime-editing guide RNAs. *Nature Biomedical Engineering 2020 5:2* 5: 190–194

Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K & Kühn R (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol*

Cicalese MP, Ferrua F, Castagnaro L, Pajno R, Barzaghi F, Giannelli S, Dionisio F, Brigida I, Bonopane M, Casiraghi M, *et al* (2016) Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency. *Blood*

Civin CI, Strauss LC, Brovall C, Fackler MJ, Schwartz JF & Shaper JH (1984) Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J Immunol*

Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, *et al* (2019) Accurate and rapid analysis of genome editing data from nucleases and base editors with CRISPResso2. *Nat Biotechnol* 37: 224

Copley MR, Beer PA & Eaves CJ (2012) Hematopoietic stem cell heterogeneity takes center stage. *Cell Stem Cell* doi:10.1016/j.stem.2012.05.006 [PREPRINT]

DeKelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijsing SH, Sancak Y, Cui X, Steine EJ, Miller JC, *et al* (2010) Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res*

Dever DP, Bak RO, Reinisch A, Camarena J, Washington G, Nicolas CE, Pavel-Dinu M, Saxena N, Wilkens AB, Mantri S, *et al* (2016) CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature*

Dever DP & Porteus MH (2017) The changing landscape of gene editing in hematopoietic stem cells: A step towards Cas9 clinical translation. *Curr Opin Hematol* doi:10.1097/MOH.0000000000000385 [PREPRINT]

DiPersio JF, Micallef IN, Stiff PJ, Bolwell BJ, Maziarz RT, Jacobsen E, Nademanee A, McCarty J, Bridger G & Calandra G (2009a) Phase III prospective randomized double-blind placebo-controlled trial of plerixafor plus granulocyte colony-

stimulating factor compared with placebo plus granulocyte colony-stimulating factor for autologous stem-cell mobilization and transplantation for patients with non-Hodgkin's lymphoma. *J Clin Oncol* 27: 4767–4773

DiPersio JF, Stadtmauer EA, Nademanee A, Micallef INM, Stiff PJ, Kaufman JL, Maziarz RT, Hosing C, Früehauf S, Horwitz M, *et al* (2009b) Plerixafor and G-CSF versus placebo and G-CSF to mobilize hematopoietic stem cells for autologous stem cell transplantation in patients with multiple myeloma. *Blood* 113: 5720–5726

Doman JL, Pandey S, Neugebauer ME, An M, Davis JR, Randolph PB, McElroy A, Gao XD, Raguram A, Richter MF, *et al* (2023) Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* 186: 3983–4002

Doman JL, Raguram A, Newby GA & Liu DR (2020) Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nature Biotechnology 2020 38:5* 38: 620–628

Doulatov S, Notta F, Laurenti E & Dick JE (2012) Hematopoiesis: A human perspective. *Cell Stem Cell* doi:10.1016/j.stem.2012.01.006 [PREPRINT]

Edraki A, Mir A, Ibraheim R, Gainetdinov I, Yoon Y, Song CQ, Cao Y, Gallant J, Xue W, Rivera-Pérez JA, *et al* (2019) A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Mol Cell* 73: 714-726.e4

Escribano-Díaz C, Orthwein A, Fradet-Turcotte A, Xing M, Young JTF, Tkáč J, Cook MA, Rosebrock AP, Munro M, Canny MD, *et al* (2013) A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. *Mol Cell*

Esrick EB, Lehmann LE, Biffi A, Achebe M, Brendel C, Ciuculescu MF, Daley H, MacKinnon B, Morris E, Federico A, *et al* (2021) Post-Transcriptional Genetic Silencing of BCL11A to Treat Sickle Cell Disease. *N Engl J Med* 384: 205–215

Essers MAG, Offner S, Blanco-Bose WE, Waibler Z, Kalinke U, Duchosal MA & Trumpp A (2009) IFNα activates dormant haematopoietic stem cells in vivo. *Nature*

Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ & Church GM (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods 2013 10:11* 10: 1116–1121

Everette KA, Newby GA, Levine RM, Mayberry K, Jang Y, Mayuranathan T, Nimmagadda N, Dempsey E, Li Y, Bhoopalan SV, *et al* (2023) Ex vivo prime editing of patient haematopoietic stem cells rescues sickle-cell disease phenotypes after engraftment in mice. *Nature Biomedical Engineering 2023 7:5* 7: 616–628

Fares I, Chagraoui J, Gareau Y, Gingras S, Ruel R, Mayotte N, Csaszar E, Knapp DJHF, Miller P, Ngom M, *et al* (2014) Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science (1979)*

Fares I, Chagraoui J, Lehnertz B, MacRae T, Mayotte N, Tomellini E, Aubert L, Roux PP & Sauvageau G (2017) EPCR expression marks UM171-expanded CD34+ cord blood stem cells. *Blood*

Ferrari S, Jacob A, Beretta S, Unali G, Albano L, Vavassori V, Cittaro D, Lazarevic D, Brombin C, Cugnata F, *et al* (2020a) Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. *Nat Biotechnol* 38: 1298–1308

Ferrari S, Jacob A, Beretta S, Unali G, Albano L, Vavassori V, Cittaro D, Lazarevic D, Brombin C, Cugnata F, *et al* (2020b) Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. *Nat Biotechnol* 38: 1298

Ferrari S, Jacob A, Cesana D, Laugel M, Beretta S, Varesi A, Unali G, Conti A, Canarutto D, Albano L, *et al* (2022a) Choice of template delivery mitigates the genotoxic risk and adverse impact of editing in human hematopoietic stem cells. *Cell Stem Cell* 29: 1428-1444.e9

Ferrari S, Jacob A, Cesana D, Laugel M, Beretta S, Varesi A, Unali G, Conti A, Canarutto D, Albano L, *et al* (2022b) Choice of template delivery mitigates the genotoxic risk and adverse impact of editing in human hematopoietic stem cells. *Cell Stem Cell* 29: 1428-1444.e9

Ferrari S, Valeri E, Conti A, Scala S, Aprile A, Di Micco R, Kajaste-Rudnitski A, Montini E, Ferrari G, Aiuti A, *et al* (2023) Genetic engineering meets hematopoietic stem cell biology for next-generation gene therapy. *Cell Stem Cell* 30: 549–570

Ferreira da Silva J, Oliveira GP, Arasa-Verge EA, Kagiou C, Moretton A, Timelthaler G, Jiricny J & Loizou JI (2022) Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nature Communications 2022 13:1* 13: 1–11

Ferrua F, Cicalese MP, Galimberti S, Giannelli S, Dionisio F, Barzaghi F, Migliavacca M, Bernardo ME, Calbi V, Assanelli AA, *et al* (2019) Lentiviral haemopoietic stem/progenitor cell gene therapy for treatment of Wiskott-Aldrich syndrome: interim results of a non-randomised, open-label, phase 1/2 clinical study. *Lancet Haematol*

Fiumara M, Ferrari S, Omer-Javed A, Beretta S, Albano L, Canarutto D, Varesi A, Gaddoni C, Brombin C, Cugnata F, *et al* (2023) Genotoxic effects of base and prime editing in human hematopoietic stem cells. *Nat Biotechnol*

Foudi A, Hochedlinger K, Van Buren D, Schindler JW, Jaenisch R, Carey V & Hock H (2009) Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat Biotechnol*

Frangoul H, Altshuler D, Cappellini MD, Chen Y-S, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, *et al* (2021) CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N Engl J Med* 384: 252–260

Fu B, Liao J, Chen S, Li W, Wang Q, Hu J, Yang F, Hsiao S, Jiang Y, Wang L, *et al* (2022) CRISPR-Cas9-mediated gene editing of the BCL11A enhancer for pediatric β0/β0 transfusion-dependent β-thalassemia. *Nat Med* 28: 1573–1580

Fu Y, Sander JD, Reyon D, Cascio VM & Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*

Garcia V, Phelps SEL, Gray S & Neale MJ (2011) Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature*

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI & Liu DR (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature 2017 551:7681* 551: 464–471

Gaudelli NM, Lam DK, Rees HA, Solá-Esteves NM, Barrera LA, Born DA, Edwards A, Gehrke JM, Lee SJ, Liquori AJ, *et al* (2020a) Directed evolution of adenine base

editors with increased activity and therapeutic application. *Nature Biotechnology 2020 38:7* 38: 892–900

Gaudelli NM, Lam DK, Rees HA, Solá-Esteves NM, Barrera LA, Born DA, Edwards A, Gehrke JM, Lee SJ, Liquori AJ, *et al* (2020b) Directed evolution of adenine base editors with increased activity and therapeutic application. *Nature Biotechnology 2020 38:7* 38: 892–900

Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, Bauer DE, Pinello L & Joung JK (2018) An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nature Biotechnology 2018 36:10* 36: 977–982

Genovese P, Schiroli G, Escobar G, Di Tomaso T, Firrito C, Calabria A, Moi D, Mazzieri R, Bonini C, Holmes MC, *et al* (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*

Gentner B, Schira G, Giustacchini A, Amendola M, Brown BD, Ponzoni M & Naldini L (2009) Stable knockdown of microRNA in vivo by lentiviral vectors. *Nat Methods*

Gentner B, Tucci F, Galimberti S, Fumagalli F, De Pellegrin M, Silvani P, Camesasca C, Pontesilli S, Darin S, Ciotti F, *et al* (2021) Hematopoietic Stem- and Progenitor-Cell Gene Therapy for Hurler Syndrome. *N Engl J Med* 385: 1929–1940

Gottlieb TM & Jackson SP (1993) The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. *Cell*

Goyama S, Wunderlich M & Mulloy JC (2015) Xenograft models for normal and malignant stem cells. *Blood*

Greiner DL, Shultz LD, Yates J, Appel MC, Perdrizet G, Hesselton RM, Schweitzer I, Beamer WG, Shultz KL, Pelsue SC, *et al* (1995) Improved engraftment of human spleen cells in NOD/LtSz-scid/scid mice as compared with C.B-17-scid/scid mice. *American Journal of Pathology*

Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, Macaulay I, Mancini E, Luis TC, Mead A, *et al* (2016) Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun* 7

Grünewald J, Miller BR, Szalay RN, Cabeceiras PK, Woodilla CJ, Holtz EJB, Petri K & Joung JK (2023) Engineered CRISPR prime editors with compact, untethered reverse transcriptases. *Nat Biotechnol* 41: 337–343

Grünewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ & Joung JK (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature 2019 569:7756* 569: 433–437

Guo C, Ma X, Gao F & Guo Y (2023) Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol* 11

Gyurkocza B & Sandmaier BM (2014) Conditioning regimens for hematopoietic cell transplantation: one size does not fit all. *Blood* 124: 344

Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K, *et al* (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *Journal of Clinical Investigation*

Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, *et al* (2003) LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science (1979)*

Hamilton BK (2018) Current approaches to prevent and treat GVHD after allogeneic stem cell transplantation. *Hematology* 2018: 228–235

Handel E-M & Cathomen T (2011) Zinc-Finger Nuclease Based Genome Surgery: Its All About Specificity. *Curr Gene Ther*

Hao QL, Thiemann FT, Petersen D, Smogorzewska EM & Crooks GM (1996) Extended long-term culture reveals a highly quiescent and primitive human hematopoietic progenitor population. *Blood*

Hardouin G, Antoniou P, Martinucci P, Felix T, Manceau S, Joseph L, Masson C, Scaramuzza S, Ferrari G, Cavazzana M, *et al* (2023) Adenine base editor–mediated correction of the common and severe IVS1-110 (G>A) β-thalassemia mutation. *Blood* 141: 1169–1179

Van der heijden T, Seidel R, Modesti M, Kanaar R, Wyman C & Dekker C (2007) Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nucleic Acids Res* 35: 5646–5657

Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, Steinfeld I, Lunstad BD, Kaiser RJ, Wilkens AB, *et al* (2015) Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature Biotechnology 2015 33:9* 33: 985–989

Hoban MD, Cost GJ, Mendel MC, Romero Z, Kaufman ML, Joglekar A V., Ho M, Lumaquin D, Gray D, Lill GR, *et al* (2015) Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood*

Hoggatt J, Singh P, Tate TA, Chou BK, Datari SR, Fukuda S, Liu L, Kharchenko P V., Schajnovitz A, Baryawno N, *et al* (2018) Rapid mobilization reveals a highly engraftable hematopoietic stem cell. *Cell* 172: 191

Hong R & He X (2023) Development of a high-fidelity Cas9-dependent adenine base editor (ABE) system for genome editing with high-fidelity Cas9 variants. *Genes Dis* 10: 705

Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, Sontheimer EJ & Thomson JA (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc Natl Acad Sci U S A* 110: 15644–15649

Hsu JY, Grünewald J, Szalay R, Shih J, Anzalone A V., Lam KC, Shen MW, Petri K, Liu DR, Keith Joung J, *et al* (2021) PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nature Communications 2021 12:1* 12: 1–6

Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, *et al* (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature 2018 556:7699* 556: 57–63

Hutt D (2018) Engraftment, Graft Failure, and Rejection. *The European Blood and Marrow Transplantation Textbook for Nurses*: 259–270

Jayavaradhan R, Pillis DM, Goodman M, Zhang F, Zhang Y, Andreassen PR & Malik P (2019) CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nat Commun*

Jiang F & Doudna JA (2017) CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* 46: 505–529

Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E & Doudna JA (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science (1979)* 351: 867–871

Jiang T, Zhang XO, Weng Z & Xue W (2022) Deletion and replacement of long genomic sequences using prime editing. *Nat Biotechnol* 40: 227–234

Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu J-L, Zhang F, *et al* (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science (1979)* 364: 292–295

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (1979)*

Kadyk LC & Hartwell LH (1992) Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. *Genetics*

Kamel-Reid S & Dick JE (1988) Engraftment of immune-deficient mice with human hematopoietic stem cells. *Science (1979)*

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, *et al* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502: 333–339

Karikó K, Muramatsu H, Ludwig J & Weissman D (2011) Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* 39

Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, Song DW, Lee KJ, Jung MH, Kim S, *et al* (2017a) In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nature Communications 2017 8:1* 8: 1–12

Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT & Liu DR (2017b) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology 2017 35:4* 35: 371–376

Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT & Liu DR (2017c) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* 35: 371–376

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z & Joung JK (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*

Kleinstiver BP, Prew MS, Tsai SQ, Topkar V V., Nguyen NT, Zheng Z, Gonzales APW, Li Z, Peterson RT, Yeh JRJ, *et al* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature 2015 523:7561* 523: 481–485

Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahante JE, Webber BR & Moriarity BS (2018) EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J* 1: 239–250

Koblan LW, Arbab M, Shen MW, Hussmann JA, Anzalone A V., Doman JL, Newby GA, Yang D, Mok B, Replogle JM, *et al* (2021) Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat Biotechnol* 39: 1414–1425

Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A & Liu DR (2018a) Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* 36: 843

Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A & Liu DR (2018b) Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* 36: 843

Kohn DB, Booth C, Kang EM, Pai SY, Shaw KL, Santilli G, Armant M, Buckland KF, Choi U, De Ravin SS, *et al* (2020) Lentiviral gene therapy for X-linked chronic granulomatous disease. *Nat Med* 26: 200–206

Kohn DB, Booth C, Shaw KL, Xu-Bayford J, Garabedian E, Trevisan V, Carbonaro-Sarracino DA, Soni K, Terrazas D, Snell K, *et al* (2021) Autologous Ex Vivo Lentiviral Gene Therapy for Adenosine Deaminase Deficiency. *New England Journal of Medicine* 384: 2002–2013

Komor AC, Kim YB, Packer MS, Zuris JA & Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533: 420

Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH & Liu DR (2017) Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* 3

Kosicki M, Tomberg K & Bradley A (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology 2018 36:8* 36: 765–771

Kunkel TA & Erie DA (2015) Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet* 49: 291–313

Kuo CY, Long JD, Campo-Fernandez B, de Oliveira S, Cooper AR, Romero Z, Hoban MD, Joglekar A V., Lill GR, Kaufman ML, *et al* (2018) Site-Specific Gene Editing of Human Hematopoietic Stem Cells for X-Linked Hyper-IgM Syndrome. *Cell Rep*

Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, Grünewald J & Joung JK (2020) CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nature Biotechnology 2020 39:1* 39: 41–46

Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, Grünewald J & Joung JK (2021a) CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* 39: 41

Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, Grünewald J & Joung JK (2021b) CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* 39: 41–46

Kuzminov A (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc Natl Acad Sci U S A* 98: 8241–8246

Kweon J, Yoon JK, Jang AH, Shin HR, See JE, Jang G, Kim J Il & Kim Y (2021) Engineered prime editors with PAM flexibility. *Molecular Therapy* 29: 2001–2007

Lam DK, Feliciano PR, Arif A, Bohnuud T, Fernandez TP, Gehrke JM, Grayson P, Lee KD, Ortega MA, Sawyer C, *et al* (2023) Improved cytosine base editors generated from TadA variants. *Nat Biotechnol* 41: 686–697

Lapidot T, Pflumio F, Doedens M, Murdoch B, Williams DE & Dick JE (1992) Cytokine stimulation of multilineage hematopoiesis from immature human cells engrafted in SCID Mice. *Science (1979)*

Lattanzi A, Camarena J, Lahiri P, Segal H, Srifa W, Vakulskas CA, Frock RL, Kenrick J, Lee C, Talbott N, *et al* (2021) Development of β-globin gene correction in human hematopoietic stem cells as a potential durable treatment for sickle cell disease. *Sci Transl Med* 13

Laurenti E & Göttgens B (2018) From haematopoietic stem cells to complex differentiation landscapes. *Nature* 553: 418–426

Lee TH & Kang TH (2019a) DNA Oxidation and Excision Repair Pathways. *Int J Mol Sci* 20

Lee TH & Kang TH (2019b) DNA Oxidation and Excision Repair Pathways. *Int J Mol Sci* 20

Leibowitz ML, Papathanasiou S, Doerfler PA, Blaine LJ, Sun L, Yao Y, Zhang CZ, Weiss MJ & Pellman D (2021) Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nature Genetics 2021 53:6* 53: 895–905

Lewinski MK & Bushman FD (2005) Retroviral DNA Integration-Mechanism and Consequences. *Adv Genet* doi:10.1016/S0065-2660(05)55005-3 [PREPRINT]

Li C, Georgakopoulou A, Newby GA, Chen PJ, Everette KA, Paschoudi K, Vlachaki E, Gil S, Anderson AK, Koob T, *et al* (2023) In vivo HSC prime editing rescues sickle cell disease in a mouse model. *Blood* 141: 2085–2099

Li GM (1999) The role of mismatch repair in DNA damage-induced apoptosis. *Oncol Res* 11: 393–400

Li GM (2007) Mechanisms and functions of DNA mismatch repair. *Cell Research 2008 18:1* 18: 85–98

Liao J, Chen S, Hsiao S, Jiang Y, Yang Y, Zhang Y, Wang X, Lai Y, Bauer DE & Wu Y (2023) Therapeutic adenine base editing of human hematopoietic stem cells. *Nat Commun* 14

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* doi:10.1038/362709a0 [PREPRINT]

Liu B, Dong X, Cheng H, Zheng C, Chen Z, Rodríguez TC, Liang SQ, Xue W & Sontheimer EJ (2022) A split prime editor with untethered reverse transcriptase and circular RNA template. *Nat Biotechnol* 40: 1388–1393

Liu J, Guo YM, Hirokawa M, Iwamoto K, Ubukawa K, Michishita Y, Fujishima N, Tagawa H, Takahashi N, Xiao W, *et al* (2012) A synthetic double-stranded RNA, poly I: C, induces a rapid apoptosis of human CD34+ cells. *Exp Hematol*

Locatelli F, Thompson AA, Kwiatkowski JL, Porter JB, Thrasher AJ, Hongeng S, Sauer MG, Thuret I, Lal A, Algeri M, *et al* (2022) Betibeglogene Autotemcel Gene Therapy for Non-β0/β0 Genotype β-Thalassemia. *N Engl J Med* 386: 415–427

Loew R, Heinz N, Hampf M, Bujard H & Gossen M (2010) Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol*

Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, Neri M, Magnani Z, Cantore A, Lo Riso P, *et al* (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods*

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, *et al* (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol*

Long SL, Morales JC, Hwang A, Wagner MW & Boothman DA (2008) DNA mismatch repair-dependent activation of c-Abl/p73alpha/GADD45alpha-mediated apoptosis. *J Biol Chem* 283: 21394–21403

Lorenz E, Uphoff D, Reid TR & Shelton E (1951) Modification of irradiation injury in mice and guinea pigs by bone marrow injections. *J Natl Cancer Inst*

Lowsky R & Messner H (2016) Mechanisms and Treatment of Graft Failure. *Thomas' Hematopoietic Cell Transplantation: Fifth Edition* 2–2: 944–958

Maier P, Von Kalle C & Laufs S (2010) Retroviral vectors for gene therapy. *Future Microbiol* doi:10.2217/fmb.10.100 [PREPRINT]

Malkova A & Ira G (2013) Break-induced replication: functions and molecular mechanism. *Curr Opin Genet Dev* 23: 271–279

Maloisel L, Fabre F & Gangloff S (2008) DNA polymerase delta is preferentially recruited during homologous recombination to promote heteroduplex DNA extension. *Mol Cell Biol* 28: 1373–1382

Marktel S, Scaramuzza S, Cicalese MP, Giglio F, Galimberti S, Lidonnici MR, Calbi V, Assanelli A, Bernardo ME, Rossi C, *et al* (2019) Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent ß-thalassemia. *Nat Med* doi:10.1038/s41591-018-0301-6 [PREPRINT]

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR & Ploegh HL (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol*

Massberg S, Schaerli P, Knezevic-Maramica I, Köllnberger M, Tubo N, Moseman EA, Huff I V., Junt T, Wagers AJ, Mazo IB, *et al* (2007) Immunosurveillance by hematopoietic progenitor cells trafficking through blood, lymph, and peripheral tissues. *Cell* 131: 994–1008

Mathis N, Allam A, Kissling L, Marquart KF, Schmidheini L, Solari C, Balázs Z, Krauthammer M & Schwank G (2023) Predicting prime editing efficiency and product purity by deep learning. *Nat Biotechnol* 41

Matthys P, Hatse S, Vermeire K, Wuyts A, Bridger G, Henson GW, De Clercq E, Billiau A & Schols D (2001) AMD3100, a Potent and Specific Antagonist of the Stromal Cell-Derived Factor-1 Chemokine Receptor CXCR4, Inhibits Autoimmune Joint Inflammation in IFN- Receptor-Deficient Mice. *The Journal of Immunology*

Maximow AA (1909) The lymphocyte as a stem cell, common to different blood elements in embryonic development and during the post-fetal life of mammals. *Folia Haematologica*

Mayuranathan T, Newby GA, Feng R, Yao Y, Mayberry KD, Lazzarotto CR, Li Y, Levine RM, Nimmagadda N, Dempsey E, *et al* (2023) Potent and uniform fetal hemoglobin induction via base editing. *Nat Genet* 55

McAuley GE, Yiu G, Chang PC, Newby GA, Campo-Fernandez B, Fitz-Gibbon ST, Wu X, Kang SHL, Garibay A, Butler J, *et al* (2023) Human T cell generation is restored in CD3δ severe combined immunodeficiency through adenine base editing. *Cell* 186: 1398-1416.e23

McCarty DM, Young SM & Samulski RJ (2004) Integration of Adeno-Associated Virus (AAV) and Recombinant AAV Vectors. *Annu Rev Genet*

McElhinny SAN, McCarville J, Ramsden DA & Snowden CM (2000) Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol*

McGrath E, Shin H, Zhang L, Phue JN, Wu WW, Shen RF, Jang YY, Revollo J & Ye Z (2019) Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nature Communications 2019 10:1* 10: 1–9

McIntosh BE, Brown ME, Duffin BM, Maufort JP, Vereide DT, Slukvin II & Thomson JA (2015) Nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) (NBSGW) mice support multilineage engraftment of human hematopoietic cells. *Stem Cell Reports* 4: 171–180

Mikkelsen NS, Hernandez SS, Jensen TI, Schneller JL & Bak RO (2023) Enrichment of transgene integrations by transient CRISPR activation of a silent reporter gene. *Mol Ther Methods Clin Dev* 29: 1–16

Milyavsky M, Gan OI, Trottier M, Komosa M, Tabach O, Notta F, Lechman E, Hermans KG, Eppert K, Konovalova Z, *et al* (2010) A Distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. *Cell Stem Cell*

Moehle EA, Rock JM, Lee YL, Jouvenot Y, DeKelver RC, Gregory PD, Urnov FD & Holmes MC (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A*

Moscou MJ & Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science (1979)*

Mueller C & Flotte TR (2008) Clinical gene therapy using recombinant adeno-associated virus vectors. *Gene Ther* 15: 858–863

Müller M, Lee CM, Gasiunas G, Davis TH, Cradick TJ, Siksnys V, Bao G, Cathomen T & Mussolino C (2016) Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* 24: 636–644

Nakamura-Ishizu A, Takizawa H & Suda T (2014) The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development*

Naldini L (1998) Lentiviruses as gene transfer agents for delivery to non-dividing cells. *Curr Opin Biotechnol*

Naldini L (2011) Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet* doi:10.1038/nrg2985 [PREPRINT]

Naldini L, Trono D & Verma IM (2016) Lentiviral vectors, two decades later. *Science (1979)* doi:10.1126/science.aah6192 [PREPRINT]

Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone A V., An M, Newby GA, Chen JC, Hsu A, *et al* (2021) Engineered pegRNAs improve prime editing efficiency. *Nature Biotechnology 2021 40:3* 40: 402–410

Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, Wilson NK, Kent DG & Göttgens B (2016) A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 128: e20–e31

Neugebauer ME, Hsu A, Arbab M, Krasnow NA, McElroy AN, Pandey S, Doman JL, Huang TP, Raguram A, Banskota S, *et al* (2023) Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat Biotechnol* 41: 673–685

Newby GA, Yen JS, Woodard KJ, Mayuranathan T, Lazzarotto CR, Li Y, Sheppard-Tillman H, Porter SN, Yao Y, Mayberry K, *et al* (2021) Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature 2021 595:7866* 595: 295–302

Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, Noda T, Abudayyeh OO, Gootenberg JS, Mori H, *et al* (2018) Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science (1979)* 361: 1259–1262

Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I & Dick JE (2011) Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science (1979)*

Nuñez JK, Chen J, Pommier GC, Cogan JZ, Replogle JM, Adriaens C, Ramadoss GN, Shi Q, Hung KL, Samelson AJ, *et al* (2021) Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* 184: 2503-2519.e17

Ochi T, Blackford AN, Coates J, Jhujh S, Mehmood S, Tamura N, Travers J, Wu Q, Draviam VM, Robinson C V., *et al* (2015) PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science (1979)*

Olsson R, Remberger M, Schaffer M, Berggren DM, Svahn BM, Mattsson J & Ringden O (2013) Graft failure in the modern era of allogeneic hematopoietic SCT. *Bone Marrow Transplant* 48: 537–543

Omer-Javed A, Pedrazzani G, Albano L, Ghaus S, Latroche C, Manzi M, Ferrari S, Fiumara M, Jacob A, Vavassori V, *et al* (2022a) Mobilization-based chemotherapy-free engraftment of gene-edited human hematopoietic stem cells. *Cell* 185: 2248-2264.e21

Omer-Javed A, Pedrazzani G, Albano L, Ghaus S, Latroche C, Manzi M, Ferrari S, Fiumara M, Jacob A, Vavassori V, *et al* (2022b) Mobilization-based chemotherapy-free engraftment of gene-edited human hematopoietic stem cells. *Cell* 185: 2248-2264.e21

Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kühlcke K, Schilz A, Kunkel H, *et al* (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med*

Pâques F & Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev*

Park SH, Cao M, Pan Y, Davis TH, Saxena L, Deshmukh H, Fu Y, Treangen T, Sheehan VA & Bao G (2022) Comprehensive analysis and accurate quantification of unintended large gene modifications induced by CRISPR-Cas9 gene editing. *Sci Adv* 8: eabo7676

Passegué E & Ernst P (2009) IFN-α wakes up sleeping hematopoietic stem cells. *Nat Med* doi:10.1038/nm0609-612 [PREPRINT]

Pavani G, Laurent M, Fabiano A, Cantelli E, Sakkal A, Corre G, Lenting PJ, Concordet JP, Toueille M, Miccio A, *et al* (2020) Ex vivo editing of human hematopoietic stem cells for erythroid expression of therapeutic proteins. *Nat Commun* 11

Petri K, Zhang W, Ma J, Schmidts A, Lee H, Horng JE, Kim DY, Kurt IC, Clement K, Hsu JY, *et al* (2022a) CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat Biotechnol* 40: 189

Petri K, Zhang W, Ma J, Schmidts A, Lee H, Horng JE, Kim DY, Kurt IC, Clement K, Hsu JY, *et al* (2022b) CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat Biotechnol* 40: 189

Publicover A, Richardson DS, Davies A, Hill KS, Hurlock C, Hutchins D, Jenner MW, Johnson PW, Lamb J, Launders H, *et al* (2013) Use of a biosimilar granulocyte colony-stimulating factor for peripheral blood stem cell mobilization: An analysis of mobilization and engraftment. *Br J Haematol*

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP & Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173–1183

Rai R, Naseem A, Vetharoy W, Steinberg Z, Thrasher AJ, Santilli G & Cavazza A (2023) An improved medium formulation for efficient ex vivo gene editing, expansion and engraftment of hematopoietic stem and progenitor cells. *Mol Ther Methods Clin Dev* 29: 58–69

Rai R, Romito M, Rivers E, Turchiano G, Blattner G, Vetharoy W, Ladon D, Andrieux G, Zhang F, Zinicola M, *et al* (2020) Targeted gene correction of human hematopoietic stem cells for the treatment of Wiskott - Aldrich Syndrome. *Nat Commun* 11

Rajagopalan N, Kagale S, Bhowmik P & Song H (2018) A Two-Step Method for Obtaining Highly Pure Cas9 Nuclease for Genome Editing, Biophysical, and Structural Studies. *Methods Protoc* 1: 1–8

Ramirez P, Rettig MP, Uy GL, Deych E, Holt MS, Ritchey JK & DiPersio JF (2009) BIO5192, a small molecule inhibitor of VLA-4, mobilizes hematopoietic stem and progenitor cells. *Blood* 114: 1340–1343

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, *et al* (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature 2015 520:7546* 520: 186–191

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, *et al* (2013) Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*

De Ravin SS, Li L, Wu X, Choi U, Allen C, Koontz S, Lee J, Theobald-Whiting N, Chu J, Garofalo M, *et al* (2017) CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. *Sci Transl Med*

De Ravin SS, Wu X, Moir S, Anaya-O'Brien S, Kwatemaa N, Littel P, Theobald N, Choi U, Su L, Marquesen M, *et al* (2016) Lentiviral hematopoietic stem cell gene therapy for X-linked severe combined immunodeficiency. *Sci Transl Med*

Rees HA & Liu DR (2018a) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* doi:10.1038/s41576-018-0059-1 [PREPRINT]

Rees HA & Liu DR (2018b) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics 2018 19:12* 19: 770–788

Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, *et al* (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature Biotechnology 2020 38:7* 38: 883–891

Río P, Navarro S, Wang W, Sánchez-Domínguez R, Pujol RM, Segovia JC, Bogliolo M, Merino E, Wu N, Salgado R, *et al* (2019a) Successful engraftment of gene-corrected

hematopoietic stem cells in non-conditioned patients with Fanconi anemia. *Nat Med* 25: 1396–1401

Río P, Navarro S, Wang W, Sánchez-Domínguez R, Pujol RM, Segovia JC, Bogliolo M, Merino E, Wu N, Salgado R, *et al* (2019b) Successful engraftment of gene-corrected hematopoietic stem cells in non-conditioned patients with Fanconi anemia. *Nature Medicine 2019 25:9* 25: 1396–1401

Román-Rodríguez FJ, Ugalde L, Álvarez L, Díez B, Ramírez MJ, Risueño C, Cortón M, Bogliolo M, Bernal S, March F, *et al* (2019) NHEJ-Mediated Repair of CRISPR-Cas9-Induced DNA Breaks Efficiently Corrects Mutations in HSPCs from Patients with Fanconi Anemia. *Cell Stem Cell*

Romero Z, Lomova A, Said S, Miggelbrink A, Kuo CY, Campo-Fernandez B, Hoban MD, Masiuk KE, Clark DN, Long J, *et al* (2019) Editing the Sickle Cell Disease Mutation in Human Hematopoietic Stem Cells: Comparison of Endonucleases and Homologous Donor Templates. *Mol Ther* 27: 1389–1406

Rongvaux A, Takizawa H, Strowig T, Willinger T, Eynon EE, Flavell RA & Manz MG (2013) Human Hemato-Lymphoid System Mice: Current Use and Future Potential for Medicine. *Annu Rev Immunol*

Rongvaux A, Willinger T, Martinek J, Strowig T, Gearty S V., Teichmann LL, Saito Y, Marches F, Halene S, Palucka AK, *et al* (2014) Development and function of human innate immune cells in a humanized mouse model. *Nat Biotechnol*

Roth SL, Malani N & Bushman FD (2011) Gammaretroviral Integration into Nucleosomal Target DNA In Vivo. *J Virol*

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

Sadler AJ & Williams BRG (2008) Interferon-inducible antiviral effectors. *Nat Rev Immunol* doi:10.1038/nri2314 [PREPRINT]

Sakurai M, Ishitsuka K, Ito R, Wilkinson AC, Kimura T, Mizutani E, Nishikii H, Sudo K, Becker HJ, Takemoto H, *et al* (2023) Chemically defined cytokine-free expansion of human haematopoietic stem cells. *Nature* 615: 127–133

San Filippo J, Sung P & Klein H (2008) Mechanism of Eukaryotic Homologous Recombination. *Annu Rev Biochem*

Scala S, Basso-Ricci L, Dionisio F, Pellin D, Giannelli S, Salerio FA, Leonardelli L, Cicalese MP, Ferrua F, Aiuti A, *et al* (2018) Dynamics of genetically engineered hematopoietic stem and progenitor cells after autologous transplantation in humans. *Nat Med* 24: 1683–1690

Schanz S, Castor D, Fischer F & Jiricny J (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc Natl Acad Sci U S A* 106: 5593–5598

Schene IF, Joore IP, Oka R, Mokry M, van Vugt AHM, van Boxtel R, van der Doef HPJ, van der Laan LJW, Verstegen MMA, van Hasselt PM, *et al* (2020) Prime editing for functional repair in patient-derived disease models. *Nature Communications 2020 11:1* 11: 1–8

Schiroli G, Conti A, Ferrari S, della Volpe L, Jacob A, Albano L, Beretta S, Calabria A, Vavassori V, Gasparini P, *et al* (2019a) Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell*

Schiroli G, Conti A, Ferrari S, della Volpe L, Jacob A, Albano L, Beretta S, Calabria A, Vavassori V, Gasparini P, *et al* (2019b) Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* 24: 551-565.e8

Schiroli G, Conti A, Ferrari S, della Volpe L, Jacob A, Albano L, Beretta S, Calabria A, Vavassori V, Gasparini P, *et al* (2019c) Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* 24: 551-565.e8

Schiroli G, Ferrari S, Conway A, Jacob A, Capo V, Albano L, Plati T, Castiello MC, Sanvito F, Gennery AR, *et al* (2017) Preclinical modeling highlights the therapeutic potential of hematopoietic stem cell gene editing for correction of SCID-X1. *Sci Transl Med*

Scully R, Panday A, Elango R & Willis NA (2019a) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol*

Scully R, Panday A, Elango R & Willis NA (2019b) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* 20: 698–714

Seita J & Weissman IL (2010) Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med* doi:10.1002/wsbm.86 [PREPRINT]

Sessa M, Lorioli L, Fumagalli F, Acquati S, Redaelli D, Baldoli C, Canale S, Lopez ID, Morena F, Calabria A, *et al* (2016) Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial. *Lancet* 388: 476–487

Shin JJ, Schröder MS, Caiado F, Wyman SK, Bray NL, Bordi M, Dewitt MA, Vu JT, Kim WT, Hockemeyer D, *et al* (2020) Controlled Cycling and Quiescence Enables Efficient HDR in Engraftment-Enriched Adult Hematopoietic Stem and Progenitor Cells. *Cell Rep* 32

Siegner SM, Ugalde L, Clemens A, Garcia-Garcia L, Bueren JA, Rio P, Karasu ME & Corn JE (2022) Adenine base editing efficiently restores the function of Fanconi anemia hematopoietic stem and progenitor cells. *Nature Communications 2022 13:1* 13: 1–15

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX & Zhang F (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science (1979)*

Snowden JA (2016) Rebooting autoimmunity with autologous HSCT. *Blood* 127: 8–10

Sobol RW, Watson DE, Nakamura J, Yakes FM, Hou E, Horton JK, Ladapo J, Van Houten B, Swenberg JA, Tindall KR, *et al* (2002) Mutations associated with base excision repair deficiency and methylation-induced genotoxic stress. *Proc Natl Acad Sci U S A* 99: 6860–6865

Soldi M, Sergi Sergi L, Unali G, Kerzel T, Cuccovillo I, Capasso P, Annoni A, Biffi M, Rancoita PMV, Cantore A, *et al* (2020) Laboratory-Scale Lentiviral Vector Production and Purification for Enhanced Ex Vivo and In Vivo Genetic Engineering. *Mol Ther Methods Clin Dev* 19: 411

Stiff PJ, Murgo AJ, Wittes RE, DeRisi MF & Clarkson BD (1983) Quantification of the peripheral blood colony forming unit-culture rise following chemotherapy Could leukocytaphereses replace bone marrow for autologous transplantation? *Transfusion (Paris)*

Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, Spratt SK, Surosky RT, Giedlin MA, Nichol G, *et al* (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *New England Journal of Medicine*

Till JE & McCulloch EA (1961) A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. *Radiat Res*

Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ & Joung JK (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*

Tucci F, Scaramuzza S, Aiuti A & Mortellaro A (2021) Update on Clinical Ex Vivo Hematopoietic Stem Cell Gene Therapy for Inherited Monogenic Diseases. *Molecular Therapy* 29: 489

Turchiano G, Andrieux G, Klermund J, Blattner G, Pennucci V, el Gaz M, Monaco G, Poddar S, Mussolino C, Cornu TI, *et al* (2021) Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell* 28: 1136-1147.e5

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD & Holmes MC (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*

Vavassori V, Ferrari S, Beretta S, Asperti C, Albano L, Annoni A, Gaddoni C, Varesi A, Soldi M, Cuomo A, *et al* (2023) Lipid Nanoparticles Allow Efficient and Harmless Ex Vivo Gene Editing of Human Hematopoietic Cells. *Blood*

Vavassori V, Mercuri E, Marcovecchio GE, Castiello MC, Schiroli G, Albano L, Margulies C, Buquicchio F, Fontana E, Beretta S, *et al* (2021) Modeling, optimization, and comparable efficacy of T cell and hematopoietic stem cell gene editing for treating hyper-IgM syndrome. *EMBO Mol Med* 13

Verma IM & Weitzman MD (2005) GENE THERAPY: Twenty-First Century Medicine. *Annu Rev Biochem*

Vigna E & Naldini L (2000) Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. *J Gene Med*

Voit RA, Hendel A, Pruett-Miller SM & Porteus MH (2014) Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res*

Walasek MA, van Os R & de Haan G (2012) Hematopoietic stem cell expansion: Challenges and opportunities. *Ann N Y Acad Sci*

Walton RT, Christie KA, Whittaker MN & Kleinstiver BP (2020) Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (1979)* 368: 290–296

Wang H & Xu X (2017) Microhomology-mediated end joining: new players join the team. *Cell Biosci* 7

Wang J, Exline CM, Declercq JJ, Llewellyn GN, Hayward SB, Li PWL, Shivak DA, Surosky RT, Gregory PD, Holmes MC, *et al* (2015) Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat Biotechnol*

Wang J, He Z, Wang G, Zhang R, Duan J, Gao P, Lei X, Qiu H, Zhang C, Zhang Y, *et al* (2022) Efficient targeted insertion of large DNA fragments without DNA donors. *Nat Methods* 19: 331–340

Wang X, Li J, Wang Y, Yang B, Wei J, Wu J, Wang R, Huang X, Chen J & Yang L (2018) Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nature Biotechnology 2018 36:10* 36: 946–949

Wang X, Yu J & Wang J (2023a) Neural Tube Defects and Folate Deficiency: Is DNA Repair Defective? *Int J Mol Sci* 24

Wang Y, Su M, Chen Y, Huang X, Ruan L, Lv Q & Li L (2023b) Research progress on the role and mechanism of DNA damage repair in germ cell development. *Front Endocrinol (Lausanne)* 14

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A, *et al* (2010) Highly efficient reprogramming to pluripotency and directed differentiation of human cells using synthetic modified mRNA. *Cell Stem Cell* 7: 618

Weissman IL & Shizuru JA (2008) The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. *Blood*

Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, *et al* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150: 264–278

Wiehe JM, Ponsaerts P, Rojewski MT, Homann JM, Greiner J, Kronawitter D, Schrezenmeier H, Hombach V, Wiesneth M, Zimmermann O, *et al* (2007) mRNA-Mediated Gene Delivery Into Human Progenitor Cells Promotes Highly Efficient Protein Expression. *J Cell Mol Med* 11: 521

Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, Offner S, Dunant CF, Eshkind L, Bockamp E, *et al* (2008) Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell*

Wu Y, Zeng J, Roscoe BP, Liu P, Yao Q, Lazzarotto CR, Clement K, Cole MA, Luk K, Baricordi C, *et al* (2019) Highly efficient therapeutic gene editing of human hematopoietic stem cells. *Nat Med* 25: 776–783

Wunderlich M, Chou FS, Sexton C, Presicce P, Chougnet CA, Aliberti J & Mulloy JC (2018) Improved multilineage human hematopoietic reconstitution and function in NSGS mice. *PLoS One*

Xu L, Wang J, Liu Y, Xie L, Su B, Mou D, Wang L, Liu T, Wang X, Zhang B, *et al* (2019) CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic Leukemia. *New England Journal of Medicine*

Yarnall MTN, Ioannidi EI, Schmitt-Ulms C, Krajeski RN, Lim J, Villiger L, Zhou W, Jiang K, Garushyants SK, Roberts N, *et al* (2023) Drag-and-drop genome insertion

of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat Biotechnol* 41: 500–512

Yin AH, Miraglia S, Zanjani ED, Almeida-Porada G, Ogawa M, Leary AG, Olweus J, Kearney J & Buck DW (1997) AC133, a novel marker for human hematopoietic stem and progenitor cells. *Blood*

Zeng J, Wu Y, Ren C, Bonanno J, Shen AH, Shea D, Gehrke JM, Clement K, Luk K, Yao Q, *et al* (2020a) Therapeutic base editing of human hematopoietic stem cells. *Nat Med*

Zeng J, Wu Y, Ren C, Bonanno J, Shen AH, Shea D, Gehrke JM, Clement K, Luk K, Yao Q, *et al* (2020b) Therapeutic base editing of human hematopoietic stem cells. *Nature Medicine 2020 26:4* 26: 535–541

Zeng J, Wu Y, Ren C, Bonanno J, Shen AH, Shea D, Gehrke JM, Clement K, Luk K, Yao Q, *et al* (2020c) Therapeutic base editing of human hematopoietic stem cells. *Nature Medicine 2020 26:4* 26: 535–541

Zhang P, Li X, Pan C, Zheng X, Hu B, Xie R, Hu J, Shang X & Yang H (2022) Single-cell RNA sequencing to track novel perspectives in HSC heterogeneity. *Stem Cell Res Ther* 13

Zhang Y, Gao S, Xia J & Liu F (2018) Hematopoietic Hierarchy – An Updated Roadmap. *Trends Cell Biol* doi:10.1016/j.tcb.2018.06.001 [PREPRINT]

Zhao D, Li J, Li S, Xin X, Hu M, Price MA, Rosser SJ, Bi C & Zhang X (2021) Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* 39: 35–40

Zhi S, Chen Y, Wu G, Wen J, Wu J, Liu Q, Li Y, Kang R, Hu S, Wang J, *et al* (2022) Dual-AAV delivering split prime editor system for in vivo genome editing. *Mol Ther* 30: 283–294