Functional Screening in human HSPCs identifies optimized protein-based enhancers of Homology Directed Repair

Juan A. Perez-Bermejo[1], Oghene Efagene[1], William M. Matern[1], Jeffrey K. Holden[1], Shaheen Kabir[1], Glen M. Chew[1], Gaia Andreoletti[1], Eniola Catton[1], Craig L. Ennis[1], Angelica Garcia[1], Trevor L. Gerstenberg[1], Kaisle A. Hill[1], Aayami Jain[1], Kristina Krassovsky[1], Cassandra D. Lalisan[1], Daniel Lord[1], B. Joy Quejarro[1], Jade Sales-Lee[1], Meet Shah[1], Brian J. Silva[1], Jason Skowronski[1], Yuri G. Strukov[1], Joshua Thomas[1], Michael Veraz[1], Twaritha Vijay[1], Kirby A. Wallace[1], Yue Yuan[1], Jane L. Grogan[1], Beeke Wienert[1], Premanjali Lahiri[1], Sebastian Treusch[1], Daniel P. Dever[1], Vanessa B. Soros[1], James R. Partridge[1], Kristen L. Seim[1]*

*Corresponding author. E-mail: scicomms@graphitebio.com
[1]Graphite Bio, Inc., 201 Haskins Way, South San Francisco, CA 94080

Abstract

Homology Directed Repair (HDR) enables precise genome editing and holds great promise in the gene therapy field. However, the implementation of HDR-based therapies is hindered by limited efficiency in comparison to methods that exploit alternative DNA repair routes, such as Non-Homologous End Joining (NHEJ). In this study, we demonstrate the development of a functional, pooled screening platform utilizing an HDR-based readout to identify protein-based reagents that improve HDR outcomes in human hematopoietic stem and progenitor cells (HSPCs), a clinically relevant cell type for gene therapy. We leveraged this screening platform to explore sequence diversity at the binding interface of the NHEJ inhibitor i53 and its target, 53BP1, and we identified optimized i53 variants that enable new intermolecular bonds and robustly increase HDR. These variants specifically reduce insertion-deletion outcomes and also synergize with a DNAPK inhibitor to increase HDR rates. When applied at manufacturing scale, the incorporation of improved variants results in a significant increase in cells with at least one repaired allele and improved HDR in long-term HSPCs subpopulations, while not increasing off-target editing or gross chromosomal rearrangements. We anticipate the pooled screening platform will enable discovery of future gene editing reagents that improve HDR outcomes, such as the i53 variants reported here.

Introduction

The discovery of CRISPR-Cas9 has transformed the landscape of gene therapy[1,2]. *Ex vivo* genetic editing of a patient's hematopoietic stem and progenitor cells (HSPCs) has the potential to cure a wide range of diseases, such as hemoglobinopathies and primary immunodeficiencies [3–5]. The premise of CRISPR-Cas9 gene therapy relies on the introduction of a targeted DNA double-strand break (DSB) and subsequent repair by innate cellular pathways. One pathway, homology-directed repair (HDR), utilizes DNA donor templates to enable precise gene correction, resulting in a scar-free conversion of alleles from disease-causing to non-pathogenic[6].

Despite the therapeutic potential of HDR-based *ex vivo* gene therapy, its clinical implementation remains a challenge for applications requiring high efficiencies of gene correction[1,7,8]. Though targeted DSBs in

HSPCs can be efficiently engineered *ex vivo*, template-guided HDR repair of DSBs competes with the dominant non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) pathways[6,8]. NHEJ and MMEJ yield a range of unwanted sequence insertions or deletions (indels) at the break site and can lead to gene disruption. Moreover, it has been observed in HSPC xenograft models that HPSC subpopulations with long term engraftment potential preferentially utilize indel-producing pathways over HDR[9]. This can result in an enrichment of indels post-transplantation and a reduction of overall rates of HDR correction *in vivo*, potentially compromising the therapeutic potential of edited cells post engraftment. One way to mitigate this risk is to increase overall HDR correction levels *ex vivo* by increasing the amount of DNA donor provided. Unfortunately, excess AAV6 or ssODN, the preferred DNA templates for HDR in HSPCs, can further induce DNA damage response (DDR) pathways and limit proliferation, yield, and engraftment potential of edited cell pools[10–12]. In order to improve the safety and efficacy of HDR-based gene therapies, there is a growing need for methods that can increase HDR editing efficiency and/or reduce the deleterious impact of DNA repair templates.

An alternative way to improve HDR editing rates is through downregulation or inhibition of key factors involved in competing NHEJ and MMEJ pathways, such as 53BP1 and DNAPKcs, respectively[8,13–15]. However, the efficacy and safety profiles of existing DNA repair-modulating approaches in HSPCs are currently poorly understood. To increase HDR repair outcomes reliably and safely in HSPCs, we sought to systematically identify, optimize, and characterize antagonists of DNA repair pathways in *ex vivo* editing platforms using highly selective, target-specific, and transient protein-based inhibitors that can be co-delivered into cells with the RNP.

In this study, we describe the design and implementation of a pooled screen to identify protein variants with optimized HDR boosting capabilities. Notably, this screen is performed in primary HSPCs and utilizes HDR as a readout, enabling direct interrogation of DNA repair modulating proteins in a functionally and physiologically relevant context. We leveraged this screen to improve the potency and robustness of i53, an engineered ubiquitin variant[13,15] that inhibits the recruitment to DSBs of 53BP1, an NHEJ-related factor, and increase HDR-based editing outcomes. Using targeted saturation mutagenesis at the i53:53BP1 binding interface, we identified i53 variants that introduce new stabilizing interactions with 53BP1 and are more potent than i53 at increasing HDR editing outcomes in HSPCs. We also provide an extensive characterization of the resulting gene editing and cell health outcomes in HSPCs using differing levels of repair template and compare the results to a separate family of HDR enhancing small molecules. Overall, our results demonstrate the utility of our pooled functional screening system to identify promising protein-based reagents for the development of safer, more efficient HDR-based gene therapies.

Results

Functional screening in HSPCs identifies candidate HDR-enhancing proteins.

To identify new protein-based reagents that enhance HDR (***Fig. S1.1***), we developed a pooled screen platform in HSPCs that uses HDR as a direct functional readout (***Fig. 1A***, ***S1.2***). In this system, HSPCs are first transduced with lentiviral libraries containing candidate protein inhibitor cDNA sequences, then edited using CRISPR-Cas9 RNPs and an AAV6 DNA donor to introduce a GFP expression cassette (UbC-GFP) via HDR at the desired locus. The GFP positive (GFP+) HSPCs fraction (i.e., HDR-positive) and the GFP negative (GFP-) HSPC fractions (i.e. HDR-negative) are isolated via fluorescence activated cell sorting (FACS) and protein variants contained within each pool are identified via next-generation sequencing.

After the development and validation of the screening system, we leveraged this platform to identify i53 variants that display improved HDR enhancing capabilities when targeting the *HBB* locus, a genetic site of therapeutic relevance for the treatment of multiple hemoglobinopathies such as sickle cell anemia and b-thalassemia[16–18]. We first screened individual and combinatorial saturation mutagenesis libraries targeting two adjacent i53 residues L67 and H68, located at the binding interface of i53 and 53BP1 (***Fig. 1B***), and identified amino acid variants that were enriched in the GFP+ population relative to i53, in particular variants that introduce a positive charge at residue 67 (***Fig. 1C***, ***S1.3***). We then iterated on the sequences of two top hits from these libraries (L67R and L67H) by varying the amino acids T12 and T14 simultaneously, which revealed changes that could further increase the HDR-based readout. These included the introduction of negative charge or, in the case of parent L67H, a positively charged histidine at position 14 (***Fig. 1D***, ***S1.4***). In all libraries tested, consistent enrichment results were observed across degenerate codons and amino acids with similar properties. Promising variants from each screen were validated by editing HSPCs transduced using smaller pooled libraries and/or individual variant transductions (***Fig. S1.5***) using the HBB-UbC-GFP AAV6 DNA donor.

Improved i53 variants identified in each screen were recombinantly produced for biophysical characterization and assessment as purified protein-based gene editing reagents. Their binding to the 53BP1 Tudor domain were evaluated using size exclusion chromatography, biolayer interferometry, and TR-FRET. All recombinant variants complexed with the 53BP1 Tudor domain and exhibited increased binding affinity relative to parental i53 (***Fig. S1.6***), suggesting a correlation between the improved affinity of the selected variants and their enrichment in the GFP+ population in our screening platform. Taking this data altogether, we chose four i53 protein variants that represent the sequence diversity observed among top performing hits (L67H, L67R, T12V.T14H.L67H, T12Y.T14E.L67R) to further characterize as protein-based HDR enhancers in HSPCs.

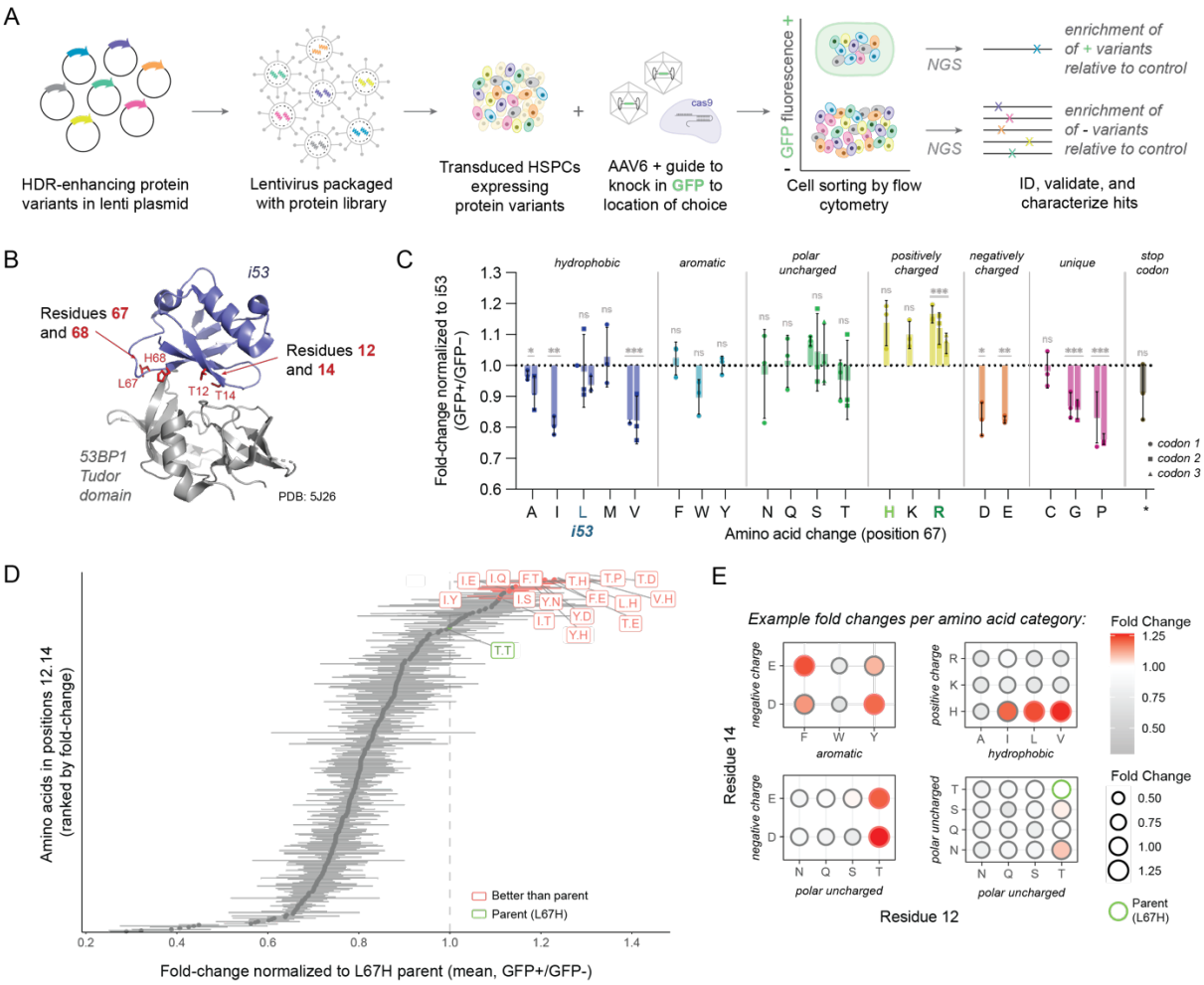


***Figure 1*: A lentiviral-based pooled screening platform in HSPCs identifies HDR-enhancing variants of i53.** (A) Schematic outlining the lentiviral-based pooled screening platform in HSPCs used to identify protein-based additives to increase HDR at a Cas9-mediated cut site of interest using an AAV6 DNA donor template. Protein variants are encoded in lentiviral libraries; once integrated into the genome the sequences can be amplified from HSPC subpopulations. To functionally quantify homology-based repair in pooled libraries, transduced cells are edited using Cas9 RNP and an AAV6 template that encodes for a GFP insertion at the cut site of interest (e.g. *HBB* gene). Post editing (3-5 days), cells are sorted via flow cytometry into GFP+ and GFP- populations. Genomic DNA is extracted from each sorted cell population, sequenced via NGS, and analyzed to determine the distribution of variants relative to a control. (B) Residues targeted for mutagenesis were chosen by their proximity to the binding interface between with 53BP1 and i53 and are shown in red (T12, T14, L67, H68). (C) Example enrichment of residues following screening with a saturation mutagenesis (NNK) library at position L67 (parent: i53). *N=3* separate pooled analyses. Each bar represents a unique codon for that amino acid. *ns = not significant; *: p-val < 0.05; **: p-val < 0.01; ***: p-val < 0.001. Two-tailed t-test with Holm-Šídák correction for multiple comparisons.* (D) Example enrichment of residues following a combinatorial library at positions T12 and T14 (parent: L67H). Variants for which all replicates were enriched over parent are highlighted in red. *N=3 separate pooled analyses.* Top hits were subsequently validated in focused libraries and experiments with purified recombinant protein. (E) Dot plot representation of variant fold change enrichment in combinatorial analysis, clustered by amino acid properties. Variations of residues 12 and 14 shown on the x-axis and y-axis, respectively.

Identified i53 variants display additional intermolecular interactions with 53BP1.

To understand the molecular basis of the improved activity and binding of the *de novo* i53 variants identified in our screens, we solved crystal structures of each variant bound to 53BP1 (***Fig. 2***, ***S2.1***; ***Table S2.1***). We observed that the L67R change enables a new 2.8 Å H-bond with D1550 of 53BP1, and that L67H forms a network of inter- and intramolecular H-bonds, including a water mediated H-bond to S1554 of 53BP1 and an H-bond to D64 on i53. The intermolecular H-bond to D64 could help lock the i53 loop conformation observed in the crystal structure and lower the entropic penalty for binding. The new interactions identified at position 67 were also present in the crystal structure for T12V.T14H.L67H and T12Y.T14E.L67R. For T12V.T14H.L67H, additional stabilizing interactions were observed at positions T12V and T14H: a water mediated bridge between T14H and Y1502 and Van der Waals contacts between T12V and 53BP1. Interestingly, the T12V mutation alters the polarity of the interface, resulting in displacement of a water molecule and formation of an intermolecular H-bond between D1521 and Y1523 in 53BP1. For the T12Y.T14E.L67R structure, T14E forms a 2.6 Å H-bond to Y1502 and a 2.9 Å H-bond to the backbone amide of Y1500. The T12Y mutation forms additional Van der Waals contacts at the protein-protein interface. These analyses demonstrate that each amino acid change in the variants selected from our screen form stabilizing interactions at the 53BP1 interface, resulting in tighter binding and validating the use of this screening approach to identify gain of function protein variants.

53BP1 antagonists increase HDR outcomes by reducing NHEJ directed repair.

Selected i53 protein variants were next assessed as HDR-boosting additives for *ex vivo* editing of HSPCs by adding the purified protein directly to the electroporation buffer containing Cas9-RNP (***Fig. S3.1***). Variants along with parental (WT) i53 were tested for HDR-mediated insertion of a fluorescent reporter at four clinically relevant loci: HBB[16], HBA1[19], CCR5[20] and IL2RG [21]. The selected i53 variants outperformed WT i53 at all loci, leading to 1.5-2.5x increase in the HDR fraction over the no protein control (***Fig. 3A***, ***S3.2***). To further evaluate the i53 variants in a clinically relevant gene editing strategy, they were used for HSPC editing using an AAV6 donor aimed at correcting the sickle cell disease causing polymorphism in the *HBB* gene (HBB-SNP AAV6), followed by amplicon sequencing as a readout. Incorporating the i53 variants resulted in a significant increase in the proportion of HDR-corrected alleles relative to WT i53 at two different protein concentrations (***Fig. 3B***). Next, we quantified the potency improvement by testing a range of concentrations for editing HSPC cells and found the variants induce a ~3-fold improvement in EC_{50} compared to parental i53 protein (***Fig. 3C***, top panel; ***S3.3***).

To further elucidate the mechanism for HDR enhancement, we characterized how the optimized i53 variants impact DNA repair outcomes in HSPCs. A detailed analysis of amplicon sequencing reads of the *HBB* locus after editing allowed us to classify the four main outcomes as follows: 1) the desired HDR event mediated by the AAV6 template (“HDR”), 2) unedited alleles (“WT”), 3) HDR mediated by the highly homologous delta-globin gene (“HBD”), and 4) indels of various lengths. To distinguish whether indels resulted from

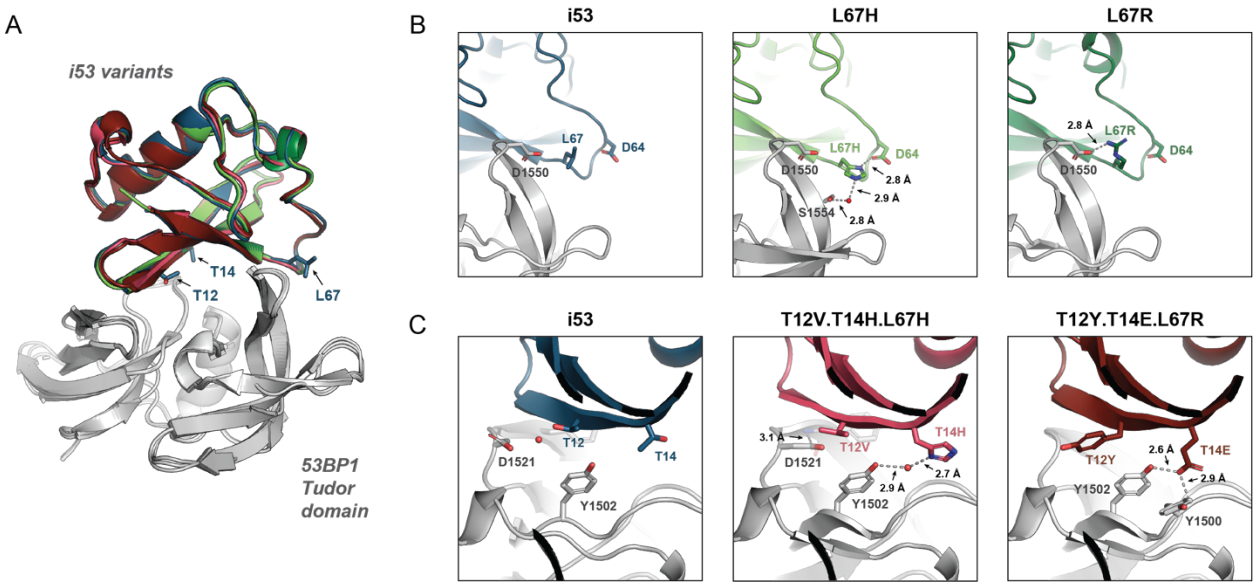


***Figure 2*: i53 variants display additional molecular interactions with 53BP1.** Crystal structures of i53 variants bound to 53BP1 Tudor domain. (A) Structural alignments of 53BP1 Tudor domain (yellow) bound to i53 variants: WT (blue, PDB: 8SVG), L67H (light green, PDB: 8SVI), L67R (dark green, PDB: 8SVH), T12V.T14H.L67H (light red, PDB: 8SVJ), and T12Y.T14E.L67R (dark red, PDB: 8T2D. (B) Zoomed in view of the complex at the solvent-exposed loop preceding β5 for i53, L67H, and L67R. (C) Zoomed in view of the complex near i53 residues 12 and 14 for wild-type i53, T12V.T14H.L67H, and T12Y.T14E.L67R. Hydrogen bonds are denoted by black dotted lines.

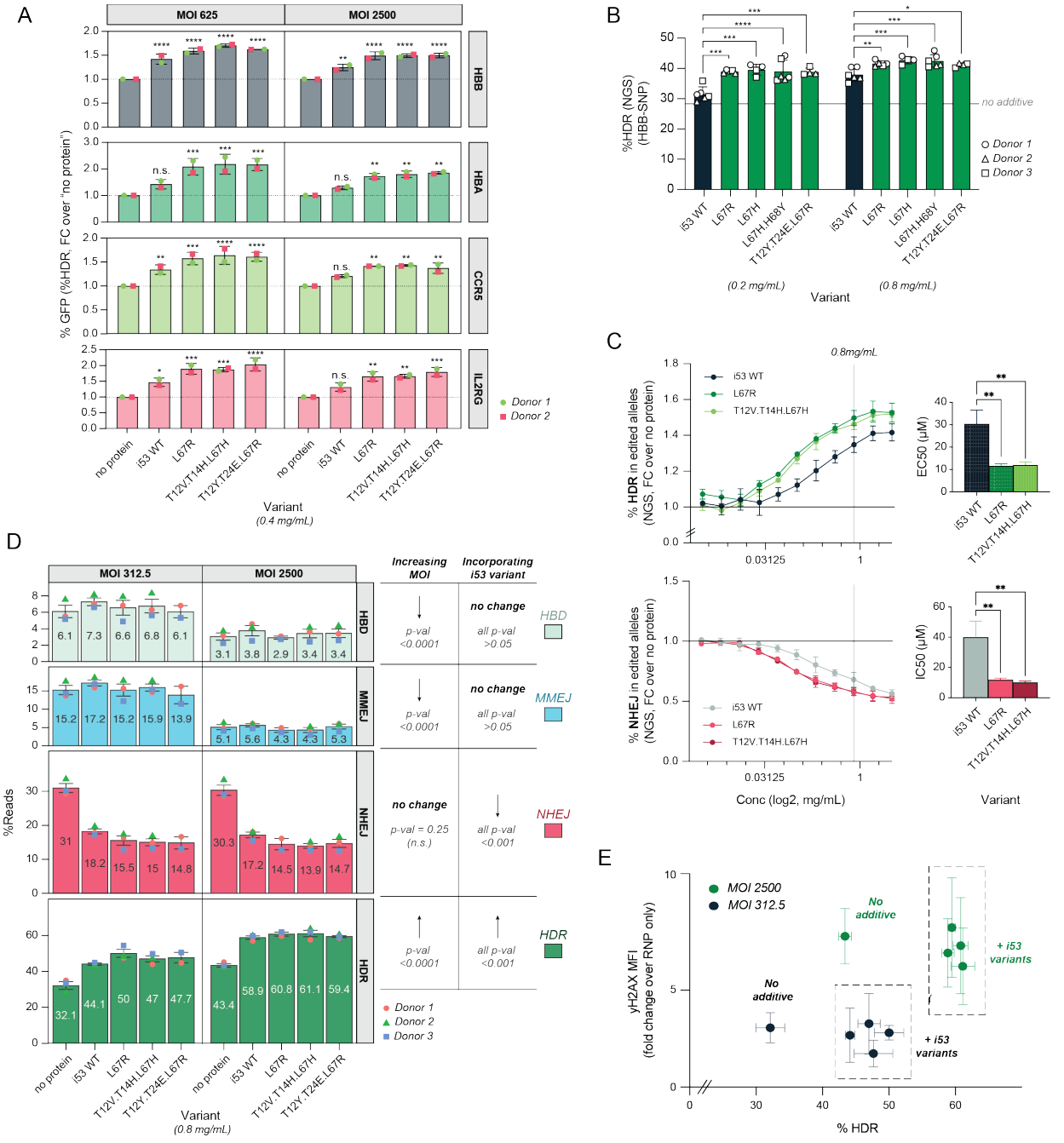


Figure 3. **i53 variants selectively target NHEJ and enable a reduction of DNA repair template.** (A) Observed fold change in %GFP-expressing cells (%HDR) when representative purified variants of i53 are incorporated as protein-based additives to HSPC editing protocols. GFP knock-ins targeted at multiple clinically relevant loci were used for readout: *HBB*, *HBA*, *CCR5*, and *Il2RG*. An equivalent plot with absolute %HDR numbers is shown in Supplemental Figure S3.2D. *N* = 2 from two different HSPC donors. Two-way ANOVA with Dunnett correction for multiple comparisons. *ns = non-significant; **: p-val < 0.01; ***: p-val < 0.001; ****: p-val < 0.0001.* (B) %HDR in HSPCs at *HBB* when edited using sickle-cell correcting HBB-SNP AAV6 and purified variants of i53 compared to parent i53 at two different protein concentrations and a MOI of 312.5. *N* = 4-6 different HSPC donors across three separate experiments. Two-way ANOVA with Dunnett correction for multiple comparisons. *ns = non-significant; **: p-val < 0.01; ***: p-val < 0.001; ****: p-val < 0.0001.* (C) Dose response curves of i53, L67R, and T12V.T14H.L67H using HBB-SNP AAV at a MOI of 625 in 3 CD34+ HSPC donors. Effects of i53 variant concentration on %HDR and %NHEJ in edited alleles are shown as fold change over no protein additive, along with the corresponding EC50s and IC50s. Vertical dotted line indicates the selected working concentration (0.8 mg/mL). *N* = 3 separate HSPC donors. Four-parameter dose response curve fit, using fold change relative to no additive (0 mg/mL); data was additionally normalized to equalize %WT readout, which shows highest variability across donors. (D) Effect of increasing MOI and incorporating i53 variants on the various editing outcomes at the HBB cut site as detected by NGS analysis. *N* = 3 different HSPC donors. Two-way ANOVA analysis, only main effects are reported. (E) %HDR relative to induction of the DNA Damage Response (DDR) as measured by flow cytometry for phosphorylation of histone H2AX (γH2AX) in cells edited with and without the addition of the i53 variants, at two different DNA repair template concentrations. *N* = 3 different HSPC donors. i53 variants tested and grouped: i53WT, L67R, T12V.T14H.L67H, T12Y.T24E.L67R. An equivalent plot for p21 expression (another DDR marker) is shown in Supplemental Figure S3.6E. p-values: MOI effect on γH2AX < 0.0001; MOI effect on HDR < 0.0001; additive effect on γH2AX < *n.s.*; additive effect on HDR < 0.0001.

MMEJ or NHEJ pathways, we used a knockdown of an enzyme known to mediate MMEJ, DNA polymerase Theta (POLQ)[22,23], and were able to further classify indels into POLQ-dependent ("MMEJ") and POLQ-independent edits (non-homologous end joining; "NHEJ") (***Fig. S3.4***). Using these classifications, we determined that i53 variants increase HDR by specifically reducing NHEJ but not MMEJ nor other types of outcomes, consistent to their role as 53BP1 antagonists (***Fig. 3C***, bottom panel; ***Fig. 3D***, ***S3.3***, ***S3.5***).

Next, we hypothesized that HDR-enhancing molecules could allow for the reduction of DNA repair template while maintaining or improving HDR outcomes, leading to a reduction in the cytotoxicity of the editing process[8,10,24]. To test this, we edited HSPCs in the presence of i53 variants with both a high and low dose of HBB-SNP AAV6 and measured cytotoxicity by yH2AX phosphorylation and p21 expression, key markers for p53-mediated DDR[7,25]. As expected, HSPC edited with i53 variants and low MOI showed comparable HDR levels to those obtained using high dose AAV6 and no additive molecule. However, we also found that the low MOI condition displayed a significant reduction in DDR marker expression (***Fig. 3E***, ***S3.6***). Interestingly, we observed that the increase in HDR associated with a higher AAV6 dose was mediated by reducing MMEJ and HBD outcomes, while NHEJ edits were unaffected. This is consistent with the observation by others that HDR displaces MMEJ indels when the homology donor is present[26]. In contrast, i53 variants increase HDR by only reducing NHEJ (***Fig. 3D***, ***S3.6***). Together, these results demonstrate that i53 variants specifically target NHEJ outcomes and can be used in combination with lower levels of AAV6 donor DNA to enhance HDR in HSPCs and mitigate DDR associated with DNA repair template.

DNAPKcs inhibitors and i53 variants can act synergistically to improve HDR

Next, we sought to study how the activity of i53 variants compares with other established NHEJ inhibitors. Small molecule inhibitors of the catalytic subunit of DNA-PK (DNAPKcs), an enzyme involved in NHEJ repair, have been previously described to increase HDR outcomes in gene editing[27–31]. We first tested a subset of commercially available DNAPK inhibitors (DNAPKi) for HDR boosting capabilities at the *HBB* locus in HSPCs and found AZD7648[31] to be the most effective (***Fig S4.1***). Comparing the use of AZD7648 with the i53 variant L67R across different loci, we found that DNAPKi and L67R have locus-dependent effects on HDR outcomes, suggesting that DNAPKcs- and 53BP1- mediated NHEJ pathways are utilized differently depending on the target genomic sequence. Indeed, we found that combining both inhibitors resulted in an improvement of HDR larger than that of L67R or AZD7648 alone (***Fig. 4A***, ***S4.2***). Thorough analysis of the editing outcomes at *HBB* revealed that, although both inhibitors contribute to a reduction of all NHEJ indels, AZD7648 preferentially reduces shorter insertions and deletions (-1, +1, +2) at the HBB locus whereas i53 variants reduce longer edits (for example -2, -5, -7). The combination of both types of inhibitors effectively reduces NHEJ indels independent of length (***Fig. 4B***, ***S4.3***).

Based on our previous observation that HDR enhancing additives can allow for a reduction of DNA template concentration while preserving HDR levels, we sought to study the impact of NHEJ inhibitors at varying

concentrations of donor template. We found that, at the locus of study *HBB*, there was an inverse correlation between AAV6 MOI and HDR fraction enrichment (compared to no additive control), although this correlation was not significant for AZD7648 alone (***Fig 4C***, ***S4.4***). Similar to i53 variant alone, use of AZD7648 alone or in combination with L67R had no significant impact on yH2AX phosphorylation or P21 expression (***Fig S4.4***).

To ascertain the safety profile of the different NHEJ inhibitors for *ex vivo* gene editing, we examined editing at a previously known off-target site for the HBB gRNA , "OT-1", by NGS[32,33]. While i53 did not impact editing at OT-1, treatment with AZD7648 resulted in a significant increase in off-target indels (~65% increase, from 1.2% to ~2%) (***Fig. 4D***). These results suggested that, even if DNA damage response markers are unaffected, addition of AZD7648 can result in an increase in off-target genotoxic effects, raising concerns about the genotoxic safety profile of AZD7648 in a clinical setting.

I53 variants increase HDR in phenotypic LT-HSCs without increasing genotoxicity.

To evaluate the feasibility of incorporating the i53 variants in a therapeutic setting HSPC cells were edited at a larger scale (200M cells per condition), closely resembling the conditions used for clinical product manufacturing[34]. Consistent with previous results, we observed that adding i53 variants increased HDR and decreased NHEJ-mediated outcomes without increasing genotoxicity (off-target editing by guide-seq and OT-1 sequencing; chromosomal translocation by karyotyping and translocation-sequencing assay) (***Fig. 5A***, ***S5.1-2***; ***Table 5.1-2***). Similarly, we observed no difference in colony forming unit recovery and composition (CFU) (***Fig S5.3***) or cell type populations by single-cell RNA sequencing (***Fig S5.4***; ***Table S5.2***) when i53 variants were included in the editing protocol.

An important metric for the ability of a genome edited HSPC cell pool to correct disease is the percent of cells (rather than alleles) that contain the desired editing outcome. It is also important to control the cells that are homozygous for detrimental outcomes such as indels, which could result in cells with no expression of the gene of interest. To address the impact of i53 variants in the single cell genotypes, we performed amplicon sequencing on individual colonies from the CFU assay. We observed that using an i53 variant resulted in a significant (~30%) increase in colonies bearing at least one HDR-corrected allele, and a significant decrease (~50%) in cells with homozygous HBB KO (indel/indel) (***Fig. 5B***, ***S5.3***).

Additionally, the success of genome edited HSPC cell therapy relies on the successful engraftment of edited cells. As such, it is important to evaluate if the HDR improvement provided by i53 variants translates to true long-term HSPCs (LT-HSPC), which is the subpopulation of CD34+ with robust engraftment potential[9,35–37]. To test this, we implemented an immunophenotyping panel to sort different subpopulations from edited HSPCs. Amplicon sequencing revealed that i53 variants provided an increase in HDR and reduction in NHEJ in all subpopulations analyzed, including LT-HSCs (***Fig. 5C***, ***S5.5***).

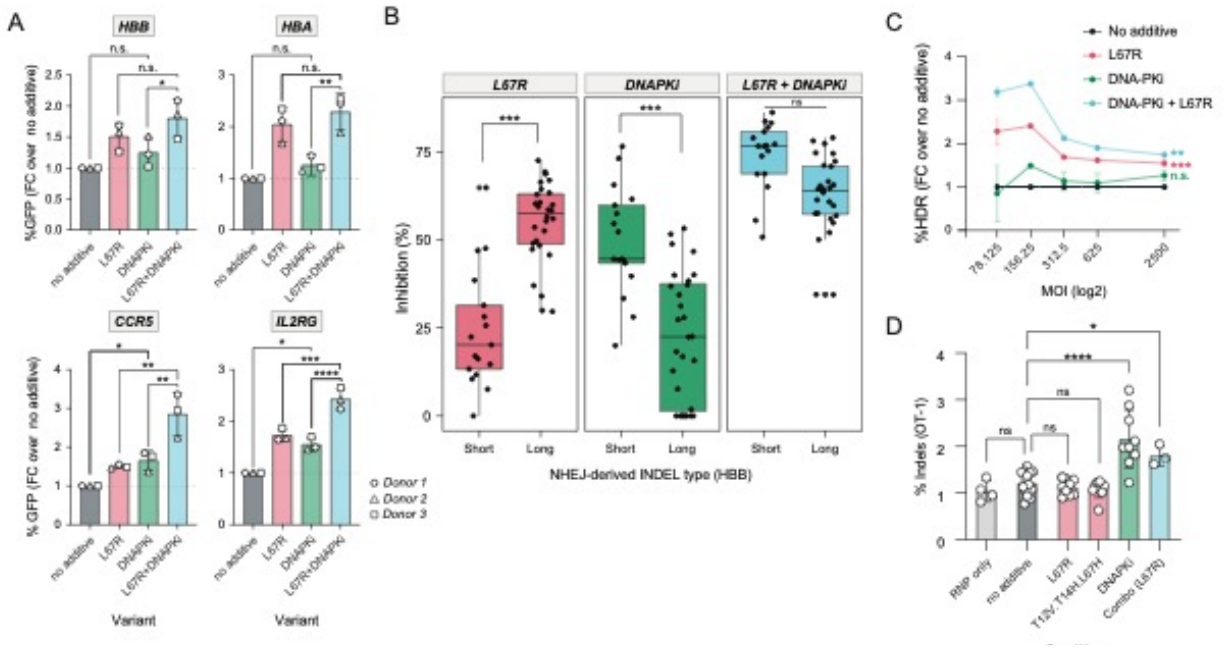


Figure 4. **i53 variants can be combined with DNAPKcs small molecule inhibitor AZD7648 for additional reduction in indels.** (A) Observed fold change in %GFP-expressing cells (%HDR) when L67R (0.8 mg/mL) is incorporated to an HSPC editing protocol for GFP knock-ins; post editing, cells were resuspended in media containing AAV targeted at *HBB, HBA, CCR5,* and *Il2RG* (MOI = 2500) with and without the addition of a DNAPKi (AZD7648, 0.5 μM). An equivalent plot with absolute %HDR numbers is shown in Supplemental Figure S4.2. *N* = 3 different donors. One-way ANOVA with Donnett adjustment for multiple comparisons. (B) Differential effects of NHEJ inhibitors (L67R and AZD7648) alone and in combination on the %inhibition of short (-1 to +3 bp in length) and long (-21 to –2 in length) NHEJ-derived INDELs at HBB. MMEJ edits are excluded. MOI = 312.5. *N = 3* different HSPC donors. Analysis by one-way ANOVA with Donnett adjustment for multiple comparisons. *ns = non-significant; ***: p-val < 0.001.* (C) %HDR in HSPCs at *HBB* when edited using sickle-cell correcting HBB-SNP AAV and L67R (0.8 mg/mL) with and without the incorporation of AZD7648 (0.5 μM), at increasing MOIs. *N* = 2 separate HSPC donors. Significance values are provided for slope being different from zero. *n.s.*: not significant (*p-value* = 0.2717), **: *p-val*< 0.01, ***: *p-val* < 0.0001. **(**D) Off target indels at OT-1 in cells edited at *HBB* without NHEJ inhibitors as compared to OT-1 indels in cells edited with the addition of i53 variants (0.8 mg/mL), AZD7648 (0.5 μM) or a combination. *N* = 4-12 different HSPC donors, compiled from 4 different experiments. Analysis by one-way ANOVA with Donnett adjustment for multiple comparisons. *ns = non-significant; *: p-val < 0.05; ****: p-val < 0.0001.*

Taken together, these results show that using i53-derived protein inhibitors such as the ones identified with our functional screening platform can drastically improve the fraction of phenotypic LT-HSCs carrying at least one HDR edited allele. To our knowledge, this improvement is provided without an increase in off-target editing or other cytotoxicity events.

Discussion

One of the most important challenges for therapeutic CRISPR-Cas9-mediated genome editing is controlling DNA repair outcomes to ensure clinically relevant levels of the desired change are introduced[1]. To identify functional protein-based additives that can modulate DNA repair and improve HDR-based outcomes in *ex vivo* gene therapy workflows, we developed a pooled screening platform that uses HDR as the functional readout. Given editing outcomes and DNA repair pathway utilization are known to be cell type dependent[38,39], we screened in primary human HSPCs, a cell type used for *ex vivo* gene therapy, enabling the identification of optimized candidates that would likely function in therapeutic applications. In this study, we leverage the functional screen to assess focused libraries targeting the protein-protein interface of i53, an engineered ubiquitin derivative, and its binding target, 53BP1. By screening in this system, we facilitated the discovery and selection of tighter binding i53 protein variants that are stable, soluble, and inherently functional for boosting HDR in HSPCs. Crystal structures of i53 variants selected through each round of screening confirmed that each i53 mutation provides new non-covalent interactions with 53BP1, resulting in increased potency through further downregulation of NHEJ outcomes.

The observation that 53BP1 antagonists exhibit a different specificity towards editing outcomes compared to DNAPKcs or MMEJ inhibition confirms that multiple orthogonal, or at least partially independent mechanisms, contribute to indel outcomes[40]. In addition, we also report that an increase in the concentration of homologous repair template (e.g., Adeno Associated Virus (AAV)) results in an enrichment of HDR outcomes by downregulation of alternative homology-mediated outcomes, especially MMEJ, possibly by competition with the endogenous repair templates. Taken together, these findings suggest a model where the sequence composition of the gene editing outcome can be modified in a locus specific manner to target specific pathways and steer the gene repair machinery towards desired outcomes, enabling more precise and predictable gene editing approaches.

Our data demonstrates that NHEJ inhibitors that increase HDR rates also enable the reduction of template DNA required for precise editing. Typically, the repair template doses required for efficient HDR in HSPCs are inherently cytotoxic, especially for LT-HSC, compromising long term stability and potential for engraftment[10–12] . We show that using NHEJ inhibitors and reducing the DNA template results in a net improvement of DDR metrics, which could improve the therapeutic potential of an edited cell pool.

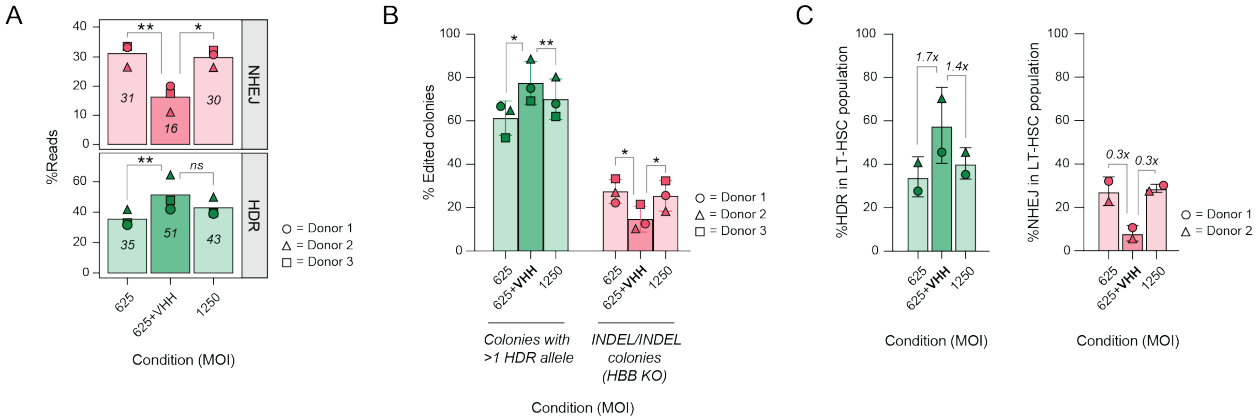


Figure 5: **i53 variants increase HDR in phenotypic LT-HSC subpopulations and increase the number of cells with successfully corrected alleles.** (A) HDR and NHEJ editing outcomes in CD34+ HSPCs that were edited with Cas9 RNP and HBB-SNP AAV6 in 3 medium scale manufacturing runs (~200 M HSPCs per condition for each donor). *N* = 3 different HSPC donors. Analysis by two-tailed paired t-test, with Holm-Šídák multiple comparison correction. *ns = non-significant (p-value = 0.1275); *: p-val < 0.05; **: p-val < 0.01.* (B) Percent of colonies bearing either ≥1 HDR allele (green) or 2 indel alleles (red). Individual colonies were genotyped at *HBB* cutsite locus. N=3 different HSPC donors. Analysis by paired t-test with Holm-Šídák multiple comparison correction. *: *p-val* < 0.05; **: *p-val* < 0.01. (C) %HDR and %NHEJ in LT-HSC-enriched (CD34+CD45RA-CD90+CD201+CD49f+CD49c+) HSPC subpopulations sorted from the bulk edited cells (donor 1 and 3 only). VHH = T12V.T14H.L67H. *N* = 2. Numbers above bars denote average fold change across two donors.

Furthermore, by using reduced AAV6 levels and an improved i53 variant to edit cells at clinically relevant scale, we validated that the HDR improvement persists in colony-forming and LT-HSC cells, significantly increasing the fraction of cells bearing repaired alleles and outlining the optimized inhibitors as promising additions to clinical *ex vivo* gene therapy workflows.

This study also highlights the importance of comprehensively characterizing any potentially detrimental effects that could impact the prospect of therapeutic gene editing. The observation that AZD7648 significantly increased indels in the main off-target site for *HBB* editing is surprising and raises concerns about the genotoxicity profile and potential for chromosomal translocations of DNAPK inhibitors for clinical gene therapy applications. Future studies will be needed to investigate whether this increase is driven by the target inhibition or specific to the specific small molecule being used. Of note, we were not able to observe any undesired effects in off-target editing, chromosomal translocation, karyotyping, or cell type composition when using the optimized 53BP1 protein antagonists for *HBB* locus editing, providing an additional level of confidence on the safety of these reagents. Despite these results, additional long-term studies using *in vivo* models will be required to further vet safety and applicability of these molecules for clinical ex-vivo cell and gene therapies. Future studies will also be needed to evaluate the efficacy and safety profile of these antagonists when used in genome editing of additional loci or in different cell types, allowing for more generalized understanding of the implications of NHEJ pathway inhibition in therapeutic genome editing.

In conclusion, we have described the development of a functional screening strategy for identifying improved 53BP1 antagonists that increase HDR repair outcomes in HSPCs. This strategy was leveraged to identify i53 variants with new stabilizing interactions with 53BP1 that improve HDR fraction enrichment and decrease NHEJ. In addition, we provide an extensive characterization of the DNA repair outcomes when these optimized i53 variants and other DNA repair pathway inhibitors are used in the editing of the *HBB locus*, a clinically relevant target for gene therapy. Notably, the screening platform described in this study can be easily tailored to identify protein-based inhibitors of other DNA repair pathways (for example, MMEJ) or to find optimized reagents for cell editing in cell types other than HSPCs. This platform can also be paired with other library selection-based methods (such as phage-, ribosome-, or yeast-display) to increase throughput and enable rapid discovery of optimized gene editing reagents. This type of screening platform presents great potential to identify functionally optimized gene editing reagents that help modulate DNA repair at specific loci, expanding the therapeutic genome editing toolkit and eventually contributing to the development of safe, efficacious, and more precise gene editing therapies.

Acknowledgements

We thank Chris Bandaro for help with DNA Damage Response (DDR) assay development and lentiviral production troubleshooting, Ryan Rodriguez for guidance on optimizing lentiviral transduction of HSPCs, and Dana Duan for assistance with lentiviral production. We also thank Ardian Wibowo, Joshua Carter, Anne Mulichak and Matt Kelker at Helix BioStructures for protein crystallization and X-ray crystallography data collection. We also would like to thank Akanksha Chhabra for her expert advice on CD34+ cell type characterizations and prioritizations.

Author contributions

Experimental design and manuscript writing: KLS and JAPB, with input from the other authors
Cloning and construct design: OE and KLS
Lenti generation and transductions: OE and KLS
Screen design, development, and implementation: OE, WMM, GMC, BJS, BW, KLS
Protein purification and biophysical analysis: JKH and JRP
Crystal Structures: JKH and JRP
Small scale HSPC editing and analysis: JAPB, OE, SK, EG, TLG, BJQ, YGS, MV, KAW, BW, KLS
Large scale HSPC editing and analysis: AG, DL, JS, MS, JSL, JT, YY, PL, JAPB
Screening NGS and analysis: JAPB, WMM, KAH, CDL, ST
HBB editing NGS and analysis: JAPB, GA, CLE, KAH, AJ, KK, CDL, TV
Project guidance: JLG, VBS, DPD

Competing Interests Statement

All authors are current or former employees of Graphite Bio, Inc. and may own stock/options in the company. A patent application encompassing aspects of this work has been filed with J.A.P.B., O.E., W.M.M., B.W., V.B.S., J.R.P. and K.L.S included as inventors and Graphite Bio, Inc. as the applicant.

Data Availability Statement

Atomic coordinates for crystal structures have been deposited in the Protein Data Bank (PDB) under accession codes: 8SVG, 8SVH, 8SVI, 8SVJ, and 8T2D. Raw and processed data files, such as those from amplicon sequencing, are available from the authors upon request.

Materials and Methods

Molecular Cloning (i53 Lentiviral Vectors)

The construct for the lentiviral-based expression and screening of i53 variants was cloned from a third-generation lentiviral plasmid (Lenti SFFV) purchased from Twist Biosciences. An empty vector was constructed to include BamHI and NsiI restriction enzyme cut sites upstream of a T2A-mCherry-WPRE cassette (to enable fluorescence-based monitoring of cells expressing the i53 variants). The sequences for i53 variants were either ordered as gene fragments from Twist Biosciences or IDT or amplified from previously constructed plasmids using primers designed to introduce the desired amino acid variation(s). Pooled NNK and combinatorial libraries were constructed using NNK primers (IDT) or oligo pools (Twist Biosciences), respectively. Combinatorial libraries were designed using one codon per amino acid (A = GCA, C = TGC, D = GAC, E = GAG, F = TTC, G = GGT, H = CAC, I = ATA, K = AAA, L = CTC, M = ATG, N = AAT, P = CCT, Q = CAG, R = CGG, S = AGC, T = ACC, V = GTC, W = TGG, Y = TAT). In certain library compositions, cysteine (C) and methionine (M) were excluded due to their inherent reactivity and susceptibility to oxidation (for any given two position combinatorial library, N = 324 - 400). Variants and libraries were cloned into the digested empty vector at the BamHI/NsiI cut sites using standard Gibson assembly protocols.

>MT_vector sequence

gtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggt
gggaggtctatataagcagagctcgtttagtgaaccggggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagt
gtggaaaatctctagcagtggcgcccgaacagggacctgaaagcgaaagggaaaccagagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcg
tcagtattaagcgggggagaattagatcgcgatgggaaaaaattcggttaaggccagggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatccct
tcagacaggatcagaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgag
ggacaattggagaagtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagagcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcctcaatgacgctgacggt
acaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttg
caccactgctgtgccttggaatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaacaagaattattggaattagataaatgggcaagtt
tgtggaattggtttaacataacaaattggctgtggtatataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacaggcccgaaggaatagaagaa
gaaggtggagagagagacagagacagatccattcgattagtgaacggatctcgacggtatcggttaacttttaaaagaaaaggggggattggggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttggc
tcccgatcgttgcgttacacacacaattactgctgatcgagtgtagccttcgaatgaaagaccccacctgtaggtttggcaagatagctgcagtaacgccattttgcaaggcatggaaaaataccaaaccaagaatagagaagttcagatcaagggcgggtacatgaaaatagctaacgttgggccaaacag
gatatctgcggtgagcagtttcggccccggcccggggccaagaacagatggtcaccgcagtttcggccccggcccgaggccaagaacagatggtccccagatatggcccaaccctcagcagtttcttaagacccatcagatgtttccaggctcccccaaggacctgaaatgaccctgcgccttatttgaatta
accaatcagcctgcttctcgcttctgttcgcgcgcttctgcttcccgagctctataaaagagctcacaacccctcactcggcgcgccagtcctccgattgactgagtcgccctgatcattgtcgatcctaccatccactcgacacacccgccaggatccccagcggccgcgccatctagagcatgcatgggagtgg
ggaaggtcgcggctccctgttgacttgtggcgatgtagaagagaacccaggtccagtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggc
acccagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtg
gtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaag
ctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacg
agctgtacaagtgagctagcgactatcatatgcttaccgtaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgag
gagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgaca
attccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttggg
ccgcctccccgcactagtgtctcgagggcccgcatcgaattcgagctcggtacctttaagaccaatgacttacaaggcagctgtagatcttagccactttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatctgctttttgcttgtactgggtctctctggttagaccagatctg
agcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagtgtggaaaatctctagcagtcctggccaacgtgagcaccgtgctgacctccaaatatcgttaa
gctggagcctgggagccggcctggccctccgccccccccacccccgcagcccacccctggtctttgaataaagtctgagtgagtggccgacagtgcccgtggagttctcgtgacctgaggtgcagggccggcgctagggacacgtccgtgcacgtgccgaggcccccctgtgcagctgcaagggacaggc
ctagccctgcaggcctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgcctcatggctgactaattttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggacgcttttttggaggccgaggcttttgcaaagatcgaacaagagacaggacct
gcaggttaattaaatttaaatcatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagc
tccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttga
gtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttg
atccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaa
aaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcatttaaatggccggcctggcgcgccgtttaaacctagatattgatagtctgatcggtcaacgtataatcgagtcctagcttttgca
aacatctatcaagagacaggatcagcaggaggctttcgcatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcgcgagtgggttacatcgaactggatctcaacagcggtaagatc
cttgagagttttcgccccgaagaacgctttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtattcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattat
gcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgattggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatgg
caacaaccttgcgtaaactattaactggcgaactacttactctagcttcccggcaacagttgatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagat
ggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaaccgattctaggtgcattggcgcagaaaaaaatgcctgatgcgacgctgcgcgtcttatactcccacatatgccagattcagca
acggatacggcttccccaacttgcccacttccatacgtgtcctccttaccagaaatttatccttaagatcccgaatcgtttaaacgcgatcgcagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgccc
attgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggac
tttcctacttggcagtacatctacgtatta

>i53v_vector (L67R) sequence

ccccacccccgcagcccacccctggtctttgaataaagtctgagtgagtggccgacagtgcccgtggagttctcgtgacctgaggtgcagggccggcgctagggacacgtccgtgcacgtgccgaggcccccctgtgcagctgcaagggacaggcctagccctgcaggcctaactccgcccatcccgcccct
aactccgcccagttccgcccattctccgcctcatggctgactaattttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggacgcttttttggaggccgaggcttttgcaaagatcgaacaagagacaggacctgcaggttaattaaatttaaatcatgtgagcaaaaggcc
agcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccg
gatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggca
gcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttt
ttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaa
acttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcatttaaatggccggcctggcgcgccgtttaaacctagatattgatagtctgatcggtcaacgtataatcgagtcctagcttttgcaaacatctatcaagagacaggatcagcaggaggctttc
gcatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcgcgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgctttccaatgatg
agcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtattcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggc
caacttacttctgacaacgattggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaaccttgcgtaaactattaactggcgaactactta
ctctagcttcccggcaacagttgatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacgggg
agtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaaccgattctaggtgcattggcgcagaaaaaaatgcctgatgcgacgctgcgcgtcttatactcccacatatgccagattcagcaacggatacggcttccccaacttgcccacttccatacgtgt
cctccttaccagaaatttatccttaagatcccgaatcgtttaaacgcgatcgcagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgcca
atagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctatta
ccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctat
ataagcagagctcgtttagtgaaccggggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagtgtggaaaatct
ctagcagtggcgcccgaacagggacctgaaagcgaaagggaaaccagagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtattaag
cgggggagaattagatcgcgatgggaaaaaattcggttaaggccagggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacagga
tcagaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgagggacaattgga
gaagtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagagcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcctcaatgacgctgacggtacaggccaga
caattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctg
tgccttggaatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaacaagaattattggaattagataaatgggcaagtttgtggaattggt
ttaacataacaaattggctgtggtatataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacaggcccgaaggaatagaagaagaaggtggag
agagagacagagacagatccattcgattagtgaacggatctcgacggtatcggttaacttttaaaagaaaaggggggattggggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttggctcccgatcgttg
cgttacacacacaattactgctgatcgagtgtagccttcgaatgaaagaccccacctgtaggtttggcaagatagctgcagtaacgccattttgcaaggcatggaaaaataccaaaccaagaatagagaagttcagatcaagggcgggtacatgaaaatagctaacgttgggccaaacaggatatctgcggt
gagcagtttcggccccggcccggggccaagaacagatggtcaccgcagtttcggccccggcccgaggccaagaacagatggtccccagatatggcccaaccctcagcagtttcttaagacccatcagatgtttccaggctcccccaaggacctgaaatgaccctgcgccttatttgaattaaccaatcagcc
tgcttctcgcttctgttcgcgcgcttctgcttcccgagctctataaaagagctcacaacccctcactcggcgcgccagtcctccgattgactgagtcgccctgatcattgtcgatcctaccatccactcgacacacccgccaggatccccagcggccgcatgttgattttcgtgaaaacccttaccgggaaaaccatca
ccctcgaggttgaaccctcggatacgatagaaaatgtaaaggccaagatccaggataaggaaggaattcctcctgatcagcagagactggcctttgctggcaaatcgctggaagatggacgtactttgtctgactacaatattctaaaggactctaaacgtcatcctctgttgagacttcgtgggagtggggaag
gtcgcggctccctgttgacttgtggcgatgtagaagagaacccaggtccagtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcaccca

gaccgccaagctgaaggtgaccaagggtggcccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgac
cgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaa
ggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgt
acaagtgagctagcgactatcatatgcttaccgtaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttg
tggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgt
ggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttgggccgcctc
cccgcactagtgtctcgagggcccgcatcgaattcgagctcggtacctttaagaccaatgacttacaaggcagctgtagatcttagccactttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatctgctttttgcttgtactgggtctctctggttagaccagatctgagcctgg
gagctctctggctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagtgtggaaaatctctagcagtcctggccaacgtgagcaccgtgctgacctccaaatatcgttaagctggag
cctgggagccggcctggccctccgcccc

Cell culture

Lenti-X HEK293T cells (Takara Bio) were cultured in DMEM (1X) + GlutaMAX-I (Gibco) supplemented with 10% FBS (Sigma). K562 cells (ATCC) were cultured in RPMI (Gibco) supplemented with 10% FBS and 1x penicillin–streptomycin (Gibco). All cells were grown in a humidified 37 °C incubator with 5% CO_2 and were passaged every 3-5 d.

Lentiviral production

Lenti-X HEK293T cells (Takara Bio) were seeded at a density of 4.5×10^6 cells per 10 cm dish 18-24 h prior to transfection. The prepared cells were co-transfected using the TransIT®-Lenti transfection reagent (Mirus Bio) with MISSION® Genomics Lentivirus Packaging Mix (Mirus Bio) and lentiviral plasmids containing i53 variants/libraries of interest. The viral supernatant was collected 48 h after transfection, passed through a 0.45-µm filter (Cytiva), flash frozen, and stored until use at -80 °C. Viral titers were measured by FACS in K562 cells and were typically ~0.5-1.5 × 10^7 TU/mL.

AAV production

The HBB-targeting AAV6 vectors HBB-SNP and HBB-UbC-GFP have been previously described[16,17]. All other AAV6 vectors were cloned into the pAAV-MCS plasmid (Agilent Technologies), which contains inverted terminal repeats (ITRs) derived from AAV2. Left and right homology arms (LHAs/RHAs) were derived from human genomic DNA to match the indicated length at the respective knock-in sites. The left and right homology arm lengths for the HBB, HBA, CCR5, and IL2RG donors were as follows: HBB LHA: 556 bp, HBB RHA: 449 bp, HBA LHA: 976 bp, HBA RHA: 879 bp, CCR5 LHA: 502 bp, CCR5 RHA: 500 bp, IL2RG LHA: 400 bp, IL2RG RHA: 414 bp. Each vector contained a UbC promoter, a CopGFP or (mCherry for HBB-UbC-mCherry), and a BGH polyA. UbC-GFP-BGH and UbC-mCherry-BGH were synthesized as gene fragments (Twist Bioscience) and cloned into pAAV with the corresponding LHA and RHA using standard Gibson Assembly protocols. The assembled LHA-UbC-GFP(/mCherry)-BGH-RHA sequences for HBB-UbC-mCherry, HBA-UbC-GFP, CCR5-UbC-GFP, and IL2RG-UbC-GFP AAV donors were as follows:

>HBB-UbC-mCherry

gtcctgtaagtattttgcatattctggagacgcaggaagagatccatctacatatcccaaagctgaattatggtagacaaaactcttccacttttagtgcatcaacttcttatttgtgtaataagaaaattgggaaaacgatcttcaatatgcttaccaagctgtgattccaaatattacgtaaatacacttgcaaaggagg
atgtttttagtagcaatttgtactgatggtatggggccaagagatatatcttagagggagggctgagggtttgaagtccaactcctaagccagtgccagaagagccaaggacaggtacggctgtcatcacttagacctcaccctgtggagccacaccctagggttggccaatctactcccaggagcagggaggg
caggagccagggctgggcataaaagtcagggcagagccatctattgcttacatttgcttctgacacaactgtgttcactagcaacctcaaacagacaccatggtgcacctgactcctgaggagaagtctgccgttactgcccattaccctgttatccctaggcctccgcgccgggttttggcgcctcccgcgggcg
ccccccctcctcacggcgagcgctgccacgtcagacgaagggcgcagcgagcgtcctgatccttccgcccggacgctcaggacagcggcccgctgctcataagactcggccttagaaccccagtatcagcagaaggacattttaggacgggacttgggtgactctagggcactggttttctttccagagagcg
gaacaggcgaggaaaagtagtcccttctcggcgattctgcggagggatctccgtggggcggtgaacgccgatgattatataaggacgcgccgggtgtggcacagctagttccgtcgcagccgggatttgggtcgcggttcttgtttgtggatcgctgtgatcgtcacttggtgagtagcgggctgctgggctggcc
ggggctttcgtggccgccgggccgctcggtgggacggaagcgtgtggagagaccgccaagggctgtagtctgggtccgcgagcaaggttgccctgaactgggggttggggggagcgcagcaaaatggcggctgttcccgagtcttgaatggaagacgcttgtgaggcgggctgtgaggtcgttgaaaca
aggtgggggggcatggtgggcggcaagaacccaaggtcttgaggccttcgctaatgcgggaaagctcttattcgggtgagatgggctggggcaccatctggggaccctgacgtgaagtttgtcactgactggagaactcggtttgtcgtctgttgcgggggcggcagttatggcggtgccgttgggcagtgcacc
cgtacctttgggagcgcgcgccctcgtcgtgtcgtgacgtcacccgttctgttggcttataatgcagggtggggccacctgccggtaggtgtgcggtaggcttttctccgtcgcaggacgcagggttcgggcctagggtaggctctcctgaatcgacaggcgccggacctctggtgaggggagggataagtgagg
cgtcagtttctttggtcggttttatgtacctatcttcttaagtagctgaagctccggttttgaactatgcgctcggggttggcgagtgtgttttgtgaagtttttttaggcaccttttgaaatgtaatcatttgggtcaatatgtaattttcagtgttagactagtaaattgtccgctaaattctggccgtttttggctttttttgttagacggtacc
gagctcttcgaaggatccatcgccaccatggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtg
accaagggtggcccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccc
tgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggccactacgac
gctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgtacaagtgactcgagggcgc
gcccgctgatcagcctcgacctgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggatt
gggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaacgtttcgcgcctgtggggcaaggtgaacgtggatgaagttggtggtgaggccctgggcaggttggtatcaaggttacaagacaggtttaaggagaccaatagaaactgggcatgtggagacagaga
agactcttgggtttctgataggcactgactctctctgcctattggtctattttcccacccttaggctgctggtggtctacccttggacccagaggttctttgagtcctttggggatctgtccactcctgatgctgttatgggcaaccctaaggtgaaggctcatggcaagaaagtgctcggtgcctttagtgatggcctggctca
cctggacaacctcaagggcacctttgccacactgagtgagctgcactgtgacaagctgcacgtggatcctgagaacttcagggtgagtctatgggacgct

>HBA-UbC-GFP

gctccagccggttccagctattgctttgtttacctgtttaaccagtatttacctagcaagtcttccatcagatagcatttggagagctgggggtgtcacagtgaaccacgacctctaggccagtgggagagtcagtcacacaaactgtgagtccatgacttggggcttagccagcacccaccaccccacgcgccacc
ccacaaccccgggtagaggagtctgaatctggagccgcccccagcccagccccgtgctttttgcgtcctggtgtttattccttcccggtgcctgtcactcaagcacactagtgactatcgccagagggaaagggagctgcaggaagcgaggctggagagcaggaggggctctgcgcagaaattcttttgagttc
ctatgggccagggcgtccgggtgcgcgcattcctctccgccccaggattgggcgaagcctcccggctcgcactcgctcgcccgtgtgttccccgatcccgctggagtcgatgcgcgtccagcgcgtgccaggccggggcgggggtgcgggctgactttctccctcgctagggacgctccggcgcccgaaag
gaaagggtggcgctgcgctccggggtgcacgagccgacagcgcccgaccccaacgggccggccccgccagcgccgctaccgccctgcccccgggcgagcgggatgggcgggagtggagtggcgggtggagggtggagacgtcctggccccgccccgcgtgcacccccaggggaggccgagc
ccgccgcccggccccgcgcaggccccgcccgggactcccctgcggtccaggccgcgccccgggctccgcgccagccaatgagcgccgcccggccgggcgtgcccccgcgccccaagcataaaccctggcgcgctcgcggcccggcactcttctggtccccacagactcagagagaacccaccgg
cctccgcgccgggttttggcgcctcccgcgggcgccccctcctcacggcgagcgctgccacgtcagacgaagggcgcagcgagcgtcctgatccttccgcccggacgctcaggacagcggcccgctgctcataagactcggccttagaaccccagtatcagcagaaggacattttaggacgggacttgg
gtgactctagggcactggttttctttccagagagcggaacaggcgaggaaaagtagtcccttctcggcgattctgcggagggatctccgtggggcggtgaacgccgatgattatataaggacgcgccgggtgtggcacagctagttccgtcgcagccgggatttgggtcgcggttcttgtttgtggatcgctgtgat
cgtcacttggtgagtagcgggctgctgggctggccggggctttcgtggccgccgggccgctcggtgggacggaagcgtgtggagagaccgccaagggctgtagtctgggtccgcgagcaaggttgccctgaactgggggttggggggagcgcagcaaaatggcggctgttcccgagtcttgaatggaaga
cgcttgtgaggcgggctgtgaggtcgttgaaacaaggtgggggggcatggtgggcggcaagaacccaaggtcttgaggccttcgctaatgcgggaaagctcttattcgggtgagatgggctggggcaccatctggggaccctgacgtgaagtttgtcactgactggagaactcggtttgtcgtctgttgcggggg
cggcagttatggcggtgccgttgggcagtgcacccgtacctttgggagcgcgcgccctcgtcgtgtcgtgacgtcacccgttctgttggcttataatgcagggtggggccacctgccggtaggtgtgcggtaggcttttctccgtcgcaggacgcagggttcgggcctagggtaggctctcctgaatcgacaggcg
ccggacctctggtgaggggagggataagtgaggcgtcagtttctttggtcggttttatgtacctatcttcttaagtagctgaagctccggttttgaactatgcgctcggggttggcgagtgtgttttgtgaagtttttttaggcaccttttgaaatgtaatcatttgggtcaatatgtaattttcagtgttagactagtaaattgtccgct
aaattctggccgtttttggctttttttgttagacggtaccgagctcttcgaaggatccatcgccaccatgcccgccatgaagatcgagtgccgcatcaccggcaccctgaacggcgtggagttcgagctggtgggcggcggagagggcacccccgagcagggccgcatgaccaacaagatgaagagcaccaa
aggcgccctgaccttcagcccctacctgctgagccacgtgatgggctacggcttctaccacttcggcacctaccccagcggctacgagaaccccttcctgcacgccatcaacaacggcggctacaccaacacccgcatcgagaagtacgaggacggcggcgtgctgcacgtgagcttcagctaccgctac
gaggccggccgcgtgatcggcgacttcaaggtggtgggcaccggcttccccgaggacagcgtgatcttcaccgacaagatcatccgcagcaacgccaccgtggagcacctgcaccccatgggcgataacgtgctggtgggcagcttcgcccgcaccttcagcctgcgcgacggcggctactacagcttc
gtggtggacagccacatgcacttcaagagcgccatccaccccagcatcctgcagaacgggggccccatgttcgccttccgccgcgtggaggagctgcacagcaacaccgagctgggcatcgtggagtaccagcacgccttcaagacccccatcgccttcgccagatctcgagtctagctcgagggcgcg
cccgctgatcagcctcgacctgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattg
ggaagacaatagcaggcatgctggggatggccatgcttcttgccccttgggcctccccccagcccctcctccccttcctgcacccgtacccccgtggtctttgaataaagtctgagtgggcggcagcctgtgtgtgcctgagtttttttccctcagcaaacgtgccaggcatgggcgtggacagcagctgggacaca

catggctagaacctctctgcagctggatagggtaggaaaaggcaggggcgggaggaggggatggaggagggaaagtggagccaccgcgaagtccagctggaaaaacgctggaccctagagtgctttgaggatgcatttgctctttcccgagttttattcccagacttttcagattcaatgcaggtttgctgaa
ataatgaatttatccatctttacgtttctgggcactctgtgccaagaactggctggctttctgcctgggacgtcactggtttcccagaggtcctcccacatatgggtggtgggtaggtcagagaagtcccactccagcatggctgcattgatccccccatcgttcccactagtctccgtaaaacctcccagatacaggcac
agtctagatgaaatcaggggtgcggggtgcaactgcaggccccaggcaattcaataggggctctactttcacccccaggtcaccccagaatgctcacacaccagacactgacgccctggggctgtcaagatcaggcgtttgtctctgggcccagctcagggcccagctcagcacccactcagctcccctga
ggctggggagcctgtcccattgcgactggagaggagagcggggccacagaggcctggctagaaggtcccttctccctggtgtgtgttttctctctgctgagcaggcttgcagtgcctggggtatca

>CCR5-UbC-GFP
gctttcatgaattcccccaacagagccaagctctccatctagtggacagggaagctagcagcaaaccttcccttcactacaaaacttcattgcttggccaaaaagagagttaattcaatgtagacatctatgtaggcaattaaaaacctattgatgtataaaacagtttgcattcatggagggcaactaaatacattc
taggactttataaaagatcactttttatttatgcacagggtggaacaagatggattatcaagtgtcaagtccaatctatgacatcaattattatacatcggagccctgccaaaaaatcaatgtgaagcaaatcgcagcccgcctcctgcctccgctctactcactggtgttcatctttggttttgtgggcaacatgctggtc
atcctcatcctgataaactgcaaaaggctgaagagcatgactgacatctacctgctcaacctggccatctctgacctgtttttccttcttactgtccccttcggcctccgcgccgggttttggcgcctcccgcgggcgcccccctcctcacggcgagcgctgccacgtcagacgaagggcgcagcgagcgtcctgat
ccttccgcccggacgctcaggacagcggcccgctgctcataagactcggccttagaaccccagtatcagcagaaggacattttaggacgggacttgggtgactctagggcactggttttctttccagagagcggaacaggcgaggaaaagtagtcccttctcggcgattctgcggagggatctccgtggggcg
gtgaacgccgatgattatataaggacgcgccgggtgtggcacagctagttccgtcgcagccgggatttgggtcgcggttcttgtttgtggatcgctgtgatcgtcacttggtgagtagcgggctgctgggctggccggggctttcgtggccgccgggccgctcggtgggacggaagcgtgtggagagaccgcca
agggctgtagtctgggtccgcgagcaaggttgccctgaactgggggttgggggggagcgcagcaaaatggcggctgttcccgagtcttgaatggaagacgcttgtgaggcgggctgtgaggtcgttgaaacaaggtgggggggcatggtgggcggcaagaacccaaggtcttgaggccttcgctaatgcggg
aaagctcttattcgggtgagatgggctggggcaccatctggggaccctgacgtgaagtttgtcactgactggagaactcggtttgtcgtctgttgcgggggcggcagttatggcggtgccgttgggcagtgcacccgtacctttgggagcgcgcgccctcgtcgtgtcgtgacgtcacccgttctgttggcttataatg
cagggtggggccacctgccggtaggtgtgcggtaggcttttctccgtcgcaggacgcagggttcgggcctagggtaggctctcctgaatcgacaggcgccggacctctggtgaggggagggataagtgaggcgtcagtttctttggtcggttttatgtacctatcttcttaagtagctgaagctccggttttgaactat
gcgctcggggttggcgagtgtgttttgtgaagttttttaggcaccttttgaaatgtaatcatttgggtcaatatgtaattttcagtgttagactagtaaattgtccgctaaattctggccgtttttggcttttttgttagacggtaccgagctcttcgaaggatccatcgccaccatgcccgccatgaagatcgagtgccgcatcac
cggcaccctgaacggcgtggagttcgagctggtgggcggcggagagggcacccccgagcagggccgcatgaccaacaagatgaagagcaccaaaggcgccctgaccttcagcccctacctgctgagccacgtgatgggctacggcttctaccacttcggcacctaccccagcggctacgagaaccc
cttcctgcacgccatcaacaacggcggctacaccaacacccgcatcgagaagtacgaggacggcggcgtgctgcacgtgagcttcagctaccgctacgaggccggccgcgtgatcggcgacttcaaggtggtgggcaccggcttccccgaggacagcgtgatcttcaccgacaagatcatccgcagca
acgccaccgtggagcacctgcaccccatgggcgataacgtgctggtgggcagcttcgcccgcaccttcagcctgcgcgacggcggctactacagcttcgtggtggacagccacatgcacttcaagagcgccatccaccccagcatcctgcagaacgggggccccatgttcgccttccgccgcgtggagga
gctgcacagcaacaccgagctgggcatcgtggagtaccagcacgccttcaagacccccatcgccttcgccagatctcgagtctagctcgagggcgcgcccgctgatcagcctcgacctgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccact
gtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggtgggctcactatgctgccgcccagtgggactttggaaatacaatgtgtca
actcttgacagggctctattttataggcttcttctctggaatcttcttcatcatcctcctgacaatcgataggtacctggctgtcgtccatgctgtgtttgctttaaaagccaggacggtcacctttggggtggtgacaagtgtgatcacttgggtggtggctgtgtttgcgtctctcccaggaatcatctttaccagatctcaaaa
agaaggtcttcattacacctgcagctctcattttccatacagtcagtatcaattctggaagaatttccagacattaaagatagtcatcttggggctggtcctgccgctgcttgtcatggtcatctgctactcgggaatcctaaaaactctgcttcggtgtcgaaatgagaagaagaggcacagggctgtgaggcttatctt
caccatcatgattgtttattttctcttctgggctccctacaa

>IL2RG-UbC-GFP
gcatggcatagaacggtgatgtcgggggtgggggttcagaacttccattatagaaggtaatgatttagaggagaaggtggttgagaatggtgctagtggtagtgaacagatccttcccaggatctaggtgggctgaggatttttgagtctgtgacactattgtatatccagctttagtttctgtttaccaccttacagcag
cacctaatctcctagaggacttagcccgtgtcacacagcacatatttgccacaccctctgtaaagccctggtttataaggttctttccaccggaagctatgacagaggaaacgtgtgggtggggaggggtagtgggtgagggacccaggttcctgacacagacagactacacccagggaatgaagagcaagc
gccatgtggcctccgcgccgggttttggcgcctcccgcgggcgcccccctcctcacggcgagcgctgccacgtcagacgaagggcgcagcgagcgtcctgatccttccgcccggacgctcaggacagcggcccgctgctcataagactcggccttagaaccccagtatcagcagaaggacattttaggac
gggacttgggtgactctagggcactggttttctttccagagagcggaacaggcgaggaaaagtagtcccttctcggcgattctgcggagggatctccgtggggcggtgaacgccgatgattatataaggacgcgccgggtgtggcacagctagttccgtcgcagccgggatttgggtcgcggttcttgtttgtggat
cgctgtgatcgtcacttggtgagtagcgggctgctgggctggccggggctttcgtggccgccgggccgctcggtgggacggaagcgtgtggagagaccgccaagggctgtagtctgggtccgcgagcaaggttgccctgaactgggggttggggggagcgcagcaaaatggcggctgttcccgagtcttga
atggaagacgcttgtgaggcgggctgtgaggtcgttgaaacaaggtgggggggcatggtgggcggcaagaacccaaggtcttgaggccttcgctaatgcgggaaagctcttattcgggtgagatgggctggggcaccatctggggaccctgacgtgaagtttgtcactgactggagaactcggtttgtcgtctgtt
gcgggggcggcagttatggcggtgccgttgggcagtgcacccgtacctttgggagcgcgcgccctcgtcgtgtcgtgacgtcacccgttctgttggcttataatgcagggtggggccacctgccggtaggtgtgcggtaggcttttctccgtcgcaggacgcagggttcgggcctagggtaggctctcctgaatcg
acaggcgccggacctctggtgaggggagggataagtgaggcgtcagtttctttggtcggttttatgtacctatcttcttaagtagctgaagctccggttttgaactatgcgctcggggttggcgagtgtgttttgtgaagttttttaggcaccttttgaaatgtaatcatttgggtcaatatgtaattttcagtgttagactagtaa
attgtccgctaaattctggccgtttttggcttttttgttagacggtaccgagctcttcgaaggatccatcgccaccatgcccgccatgaagatcgagtgccgcatcaccggcaccctgaacggcgtggagttcgagctggtgggcggcggagagggcacccccgagcagggccgcatgaccaacaagatgaag
agcaccaaaggcgccctgaccttcagcccctacctgctgagccacgtgatgggctacggcttctaccacttcggcacctaccccagcggctacgagaacccccttcctgcacgccatcaacaacggcggctacaccaacacccgcatcgagaagtacgaggacggcggcgtgctgcacgtgagcttcagct
accgctacgaggccggccgcgtgatcggcgacttcaaggtggtgggcaccggcttccccgaggacagcgtgatcttcaccgacaagatcatccgcagcaacgccaccgtggagcacctgcaccccatgggcgataacgtgctggtgggcagcttcgcccgcaccttcagcctgcgcgacggcggctact
acagcttcgtggtggacagccacatgcacttcaagagcgccatccaccccagcatcctgcagaacgggggccccatgttcgccttccgccgcgtggaggagctgcacagcaacaccgagctgggcatcgtggagtaccagcacgccttcaagacccccatcgccttcgccagatctcgagtctagctcga
gggcgcgcccgctgatcagcctcgacctgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaaggggg
aggattgggaagacaatagcaggcatgctggggatgcggtgggctcctgcaggtgaagccatcattaccattcacatccctcttattcctgcagctgcccctgctgggagtggggctgaacacgacaattctgacgcccaatgggaatgaagacaccacagctggtgggaaatctgggactggagggggct
ggtgagaagggtggctgtgggaagggggccgtacagagatctggtgcctgccactggccattacaatcatgtggcagaattgaaaagtggagtgggaagggcaaggggggagggttccctgcctcacgctacttcttctttctttcttgtttgtttgtttctttctttcttttgaggcagggtctcactatgttgcctaggctg
gtctcaaactcctggcctctagtgatcctcctgcctcagcctttcaaagcaccaggattacagacatgagcca

The NPM1-GFP AAV6 vector was designed using the sequence of a donor plasmid described by the Allen Institute for Cell Science(Roberts et al., 2017) which attaches an mEGFP tag to the C-terminus of NPM. LHA-linker-mEGFP-BGH-RHA was synthesized as a gene fragment (Azenta/Genewiz) and cloned into pAAV using standard Gibson Assembly protocols.

The HBB-SNP AAV6 was produced by Viralgen. The HBA-UbC-GFP AAV6 was produced by Packgene. HBB-UbC-GFP AAV6, CCR5-UbC-GFP AAV6, IL2RG-UbC-GFP AAV6 and NPM1-GFP AAV6 were produced by Vigene. Titers used for CD34+ HSPC editing experiments were determined using droplet digital PCR (ddPCR).

CD34+ HSPCs culture

Human CD34+ HSPCs were cultured as previously described[16,17]. CD34+ HSPCs were purchased from AllCells and were isolated from G-CSF-mobilized peripheral blood from healthy donors. CD34+ HSPCs were cultured at 2.5×10^5–5×10^5 cells/mL in StemSpan™-AOF (Stemcell) supplemented with stem cell factor (SCF) (100 ng/mL), thrombopoietin (TPO) (100 ng/mL), FLT3–ligand (100 ng/mL), IL-6 (100 ng/mL) (all Peprotech) and UM171 (35 nM) (Selleckchem). Cells were cultured at 37°C, 5% CO_2, and 5% O_2.

Lentiviral transduction of CD34+ HSPCs

CD34+ HSPC cells were transduced using lentivirus at MOIs of 0.25 - 1 at day 1 post thaw. Cells were concentrated using centrifugation (180 g x 7 min), counted, and added at a concentration of 4×10^6 cells/mL to media containing lentivirus, cyclosporin A (5uM, Sigma Aldrich), and Synperonic F108 (0.5 mg/mL, Sigma Aldrich). After 4 h of incubation, cells were spun down, washed once with media, and seeded into lentivirus-free media at a density of 3.5×10^5 cells/mL.

Genome editing of CD34+ HSPCs (AAV6 donor)

Chemically-modified single guide RNAs (sgRNAs) used to edit CD34+ HSPCs were purchased from Synthego. The sgRNA sequences were modified by adding 2'-O-methyl-3'-phosphorothioate at the three terminal nucleotides of the 5' and 3' ends. The target sequence for the sgRNAs used are as follows: HBB: 5'-CTTGCCCCACAGGGCAGTAA-3', HBA: 5'-GGCAAGAAGCATGGCCACCG-3', CCR5: 5'-

GCAGCATAGTGAGCCCAGAA-3', IL2RG: 5'- TGGTAATGATGGCTTCAACA-3', and NPM1: 5'-TCCAGGCTATTCAAGATCTC-3'. Cas9 protein (SpyFi Cas9) was purchased from Aldevron. The RNPs were complexed at a Cas9: sgRNA molar ratio of 1:2.5 at 25 °C for 10-15 min prior to electroporation. 48-72 h post thaw, CD34+ HSPC cells were collected, counted, and pelleted at 180 g x 7 min. The cell pellets were resuspended in Maxcyte buffer with complexed RNPs and electroporated using a Maxcyte ExPERT ATx Nucleofector. After electroporation cells were plated at 3.5-5.0x10^5 cells/mL in media supplemented with cytokines and the desired AAV6 donor added at 5.0×10^2 – 2.5×10^4 vector genomes/cell. 24 h after nucleofection, cells were spun down, washed once with media, and seeded into AAV-free media at a density of 3.5×10^5 cells/mL. Cells were harvested 2 days post nucleofection for NGS analysis or 3-5 days post nucleofection for GFP expression analysis.

When editing using purified i53 variant proteins, the proteins were added to CD34+ HSPCs cells as part of the nucleofection mix at concentrations of 0.0125 - 1.6 mg/mL (volume of added protein ≤ 1/10 of Maxcyte cuvette volume) prior to nucleofection. For editing with a DNAPK small molecule inhibitor (AZD7648, CC-115, and M314/nedisertib from Selleck Chemicals, or BAY8400 from MedChem Express), nucleofected cells were added to media containing both AAV6 and the DNAPKi at various concentrations. 24 hours after nucleofection, cells were spun down, washed with media, and seeded into AAV6 and DNAPKi-free media at a density of 3.5-5.0×10^5 cells/mL.

Screening and sorting of pooled libraries

Lentiviral-based i53 variant libraries were transduced at an MOI of ~0.2 – 0.5 (aiming for ~30% transduction and a coverage of >500 cells per library member in mCherry+/GFP+ cell population for each replicate tested). Three days after transduction, cells were edited in triplicate or quadruplet at HBB (or NPM1) as described above, using HBB-UbC-GFP donor AAV6 (or NPM1-GFP donor AAV6) at an MOI of 2.5×10^4 vector genomes/cell. Three days post editing, cells were pelleted and resuspended in media with DAPI (Miltenyi Biotec). Single, live, mCherry+/GFP+ and mCherry+/GFP- cells were collected using a FACSAria cell sorter (Becton Dickinson); purity of populations was confirmed by post-sort purity checks. Post sort, genomic DNA was harvested from each sorted cell population using a Quick-DNA 96 Plus Kit (Zymo Research). The DNA concentration of each sample was measured using a Qubit 1X dsDNA BR assay kit (ThermoFisher).

Next Generation Sequencing (NGS) of pooled libraries

An amplicon sequencing workflow was designed to sequence and quantify i53 variants within starting and post selection pools. Primers and PCR conditions were optimized to specifically amplify the entire variant coding sequence from plasmids, lentiviral libraries, as well as genomic DNA carrying lentiviral vector insertions. After the initial amplification, the i53 amplicons undergo an additional PCR amplification to add sequencing adapters and sample indexes to enable sample multiplexing. The resulting sequencing libraries were then sequenced on an Illumina MiSeq instrument using paired end reads to cover the full length of the i53 coding sequence.

NGS analysis of pooled libraries

The frequencies of i53 variants were quantified by counting the number of each observed sequence in the NGS data and then removing all unexpected sequence (ie using a prespecified “whitelist” of variants known to be contained in the pool). Spike-in tests using individual variants demonstrated the sequencing and analysis workflows could correctly estimate the frequencies of different i53 versions. This approach was used to confirm sequence diversity in plasmid and lentiviral libraries prior to screening. For quality control of screening data, key measures we considered were: the number of mapped reads (>1e5 reads per sample), the percent reads carried over from parent, and the diversity of observed sequences (i.e. minimal skewing). Fold-change enrichment of a variant was calculated by dividing normalized variant frequency in GFP+ cells by the frequencies in GFP- cells sorted from the same parent. All datasets contained an internal control (NNK-generated parent sequence) that was used to perform a last quality control of datasets, excluding sequencing runs where internal control abundance was more than 10% different from that of parent carry over control.

Data processing and visualizations were generated using R (v4.1.2) and the *ggplot2* package. Variants were ranked by fold change over parent and any variant for which either average or every replicate was over 1.0 was flagged as 'Better than parent', as highlighted in figures. Hits were ranked by average fold change and top selected candidate variants were moved forward for validation in targeted libraries, as described below.

Validation of hits via lentiviral expression

Sequences of individual i53 variants of interest were cloned into the lentiviral-based expression plasmid described above. Hits were validated either as pooled "validation libraries" (variant and control plasmids manually mixed to generate a pool of 5-25 variants) or individually. Lentivirus generated from these plasmids was used to transduce CD34+ HSPC cells at MOIs of 0.5 - 1 at day 1 post thaw. At day 4, the transduced cells were edited with HBB-UbC-GFP AAV6 at concentrations of 1.25 - 2.5×10^4 vector genomes/cell. Cells transduced with pooled validation libraries were edited in triplicate or quadruplet; cells transduced with individual variants were edited in duplicate.

For individual testing of variants, rates of integration of the HBB-UbC-GFP donor were measured using a Beckman Coulter CytoFLEX. DAPI (Miltenyi Biotec) was used to discriminate live and dead cells. mCherry expression was used to differentiate transduced cells from untransduced cells and rates of GFP integration were compared between the two populations to quantify the impact of lentiviral-based variant expression on HDR rates. Flow cytometry data were analyzed using FlowJo 10 software.For pooled validation libraries, cells were sorted and analyzed as described above. NGS analysis of the gDNA purified from sorted mCherry+GFP+ and mCherry+GFP- populations was used to determine differential variant enrichment and validate the impact of individual variants on HDR rates relative to a control.

i53 variant protein production and purification

The sequences of different i53 variants were cloned into bacterial expression plasmids, resulting in a N-terminal His-tagged fusion protein with a protease cleavage site in between the 6x-His-tag and i53 variant sequence. The resulting plasmids were transformed into *E. coli* BL21 (DE3)-RIL for protein expression. Cells were grown at 37 °C in Luria-Bertani broth supplemented with 0.4% glucose to OD600 = 0.8 and induced with 0.4 mM IPTG at 18°C for 18 hours. Cells were harvested by centrifugation, resuspended in 50 mM potassium phosphate pH 8.0, 500 mM NaCl, 20 mM imidazole, and 3 mM β-mercaptoethanol. Cells were lysed using an microfluidizer (Microfluidics). The crude lysate was immediately supplemented with 0.2 mM phenylmethylsulfonyl fluoride (PMSF) and centrifuged at 14,000g for 30 minutes. The soluble fraction was subsequently incubated with 2 ml Ni-NTA (GE Healthcare) per 1,000 ODs for 1 hour at 4 °C. Following incubation with the Ni-NTA resin, lysate was removed by pelleting the resin at 2,500 g for 3 minutes and washed 3 times with 9 bed volumes of 50 mM potassium phosphate pH 8.0, 500 mM NaCl, 20 mM imidazole, and 3 mM β-mercaptoethanol. Following the batch wash Ni-NTA resin was loaded onto a gravity column and His-tagged i53 variant protein was eluted with 6 bed volumes of 50 mM potassium phosphate pH 8.0, 300 mM NaCl, 500 mM imidazole, and 3 mM β-mercaptoethanol. Eluted protein was dialyzed overnight against 10 mM Tris/HCl pH 8.0, 200 mM NaCl, and 1 mM DTT and the 6xHis-tag was cleaved with protease. The protein was purified by anion exchange chromatography on a HiTrapQ column (GE Healthcare) via a linear NaCl gradient and twice by size exclusion chromatography using a Superdex S200 26/60 column (GE Healthcare) run in 10 mM Tris/HCl pH 8.0, 200 mM NaCl, 1 mM DTT. Proteins were concentrated to ~20 mg/mL and flash frozen for storage.

Size exclusion chromatography

Recombinantly purified 53BP1 Tudor domain (53BP1 residues 1484-1603) was mixed with recombinantly purified i53 variants at a concentration of 0.5 mg/mL each. Proteins were incubated for 30 minutes at room temperature prior to injection onto an HPLC (Agilent, 1260 Infinity II). 5 µL of protein complex was injected onto a MAbPac 4 x 300 mm SEC column with 5 µm particle size and 300 Å pore size. The HPLC was run at 0.2 mL per minute using PBS as the mobile phase and continuously measuring the absorbance at 280 nm for approximately 1 full column volume. 53BP1 Tudor domain alone has a retention time of 14.6 min.

i53 variants have a retention time of ~15.5 min. A stable complex of 53BP1 Tudor domain and i53 variants were found to have a retention time of 14.3 min.

Bio-layer interferometry (BLI)

Data were collected using an Octet R8 system (Sartorius). Purified 53BP1 Tudor domain was labeled at exposed primary amine groups with NHS-biotin using ChromaLINK NHS-Biotin protein labeling kit (Vector Laboratories). 1 equivalent of chromalink biotin was incubated with the 53BP1 Tudor domain for 2 h and buffer exchanged into fresh PBS. Labeling efficiency was calculated to be approximately 1 biotin per molecule of 53BP1 Tudor domain. Octet SA Biosensor tips (Sartorius) were incubated with biotin-labeled 53BP1 Tudor domain (ligand) for 60-80 s. The labeled tip was then dipped in 1x binding buffer (Sartorius) for 60 seconds to remove excess ligand and achieve baseline. Labeled tips were introduced to the i53 variant (analyte) for 500-600 s and the response was continuously monitored to detect association. A range of analyte concentrations were tested from a highest to lowest concentration in nM (i.e. 200, 100, 50, 25, 12.5, 6.25, 3.125). The tips were then introduced to 1x binding buffer for 5 minutes and the response was continuously monitored to detect dissociation. A dissociation constant (K_D) was calculated using a 1:1 binding model and the on-rate (k_a) and off-rate (k_d) were calculated as a change in response (nm) over time (s).

TR-FRET

Assay volumes of 20 uL (N = 4) were composed of 0.5 uM His-tagged i53, 0.5 uM c-terminal avi tagged 53BP1, 5 nM Europium labeled anti-His antibody (Perkin Elmer), 0.5x Streptavidin-xl665 (Cis Bio) and an i53 variant at concentrations ranging from 5000 nM to 4.8 nM. All assay components were prepared in a buffer composed of 50 mM Tris pH 7.5, 150 mM NaCl, 0.02% (v/v) Tween-20, and 0.05% (w/v) BSA. Each assay was incubated for 2 hrs at room temperature in a 384 well white optiplate. TR-FRET was measured on a Clariostar Plus plate reader (BMG LabTec) using the TR-FRET mode.

Crystallography

The human i53:53BP1 complex, purified in 10 mM Tris 8.0, 200 mM NaCl and 1 mM DTT was screened for crystallization at room temperature using a protein concentration of 30 mg/mL with the previously published condition[15] 0.1 M MES (2-(N-morpholino)ethanesulfonic acid) pH 6.0, 0.2 M trimethylamine N-oxide and 25% (w/v) PEG MME (polyethylene glycol monomethyl ether) 2000. Crystals grew within 7 days at 23 °C using the sitting drop vapor diffusion method. Crystals were cryoprotected by adding glycerol, 20% (v/v) final concentration, to the reservoir solution before flash-freezing in liquid nitrogen. The i53:53BP1 complex was crystallized in the $P2_12_12_1$ space group with one i53:53BP1 complex molecule per asymmetric unit cell.

Structure Determination

Structures of human i53:53BP1 Tudor domain (WT, L67H, L67R, T12Y.T14E.L67R, T12V.T14H.L67H) were solved using molecular replacement and previously published structure of WT i53:53BP1 Tudor domain (PDB code: 5J26). The final models for human i53:53BP1 Tudor domain (WT, L67H, L67R, T12Y.T14E.L67R, T12V.T14H.L67H) were built with native data and refined to an extended resolution below 1.8 Å for each dataset. All models of i53:53BP1 complex were built using COOT[41] and further refinement was completed using Refmac[42].

LC-MS

Samples of purified proteins (20 µg) were analyzed by LC-MS using a Poroshell 300SB-C8 2.1X7.5mm column coupled to an Agilent 6224 ToF (JadeBio, San Diego, CA).

Measuring targeted integration of HBB-UbC-GFP, HBA-UbC-GFP, CCR5-UbC-GFP, or IL2RG-UbC-GFP (flow cytometry-based analysis)

Rates of targeted integration of the HBB-UbC-GFP, HBA-UbC-GFP, CCR5-UbC-GFP, and IL2RG-UbC-GFP donors were measured using a Beckman Coulter CytoFLEX. DAPI (Miltenyi Biotec) was used to discriminate live and dead cells. Flow cytometry data were analyzed using FlowJo 10 software.

Measuring targeted integration of HBB-SNP (NGS-based analysis)

The frequency of homology directed repair (HDR) and other editing outcomes at HBB were measured using Next Generation Sequencing (NGS). An NGS assay was developed to determine the frequency of various sequence changes at the HBB locus by quantifying the number of alleles that have been either: (1) not edited (% WT), (2) changed by HDR to incorporate sequence differences present in the AAV repair template (% HR), or (3) mutated during the genome correction process resulting in a gene that produces mutant β-globin (% INDELs).

For this assay, genomic DNA was harvested from cells using a Quick-DNA 96 Plus Kit (Zymo Research). The DNA concentration was measured using a Qubit 1X dsDNA BR assay kit (ThermoFisher). Purified genomic DNA was then used to amplify the HBB locus via polymerase chain reaction (PCR). The PCR products were diluted using nuclease-free water to serve as the template DNA for targeted NGS library prep. An Aglient Tapestation was used to confirm the PCR product for each sample was the expected size (1410bp). A second PCR with primers carrying partial Illumina adapters was performed to amplify a 142 base pair sequence that includes the region of the HBB locus that is to be corrected during the genome correction process. The PCR products were diluted again to serve as templates in a third PCR reaction using Nextera XT index primers. This third PCR reaction was used to assign unique identifiers to each sample and to add the full length adapter sequences necessary for Illumina sequencing. The size of the PCR products was assessed on an Agilent BioAnalyzer. PCR products were then pooled, purified using a Qiagen PCR purification kit, and quantified using PicoGreen in order to ready the PCR products for sequencing.

Based on the PicoGreen concentration, the library of pooled PCR products was diluted to a final concentration of 4 nM. Sequencing was performed on a MiSeq system using an Illumina MiSeq sequencing reagent kit (V2, 300 cycles). A 10% PhiX control library was added to the sample library to improve sequence diversity and to allow for error rate measurements. The library was denatured and loaded at 8–12 pM onto the sequencing reagent cartridge. The sequencing entails paired-end 150 base pair reads and dual indexing reads. The sequencing data was demultiplexed based on the sample indexes provided and FASTQ files for each sample were generated. The FASTQ files were processed using the CRISPResso2 pipeline (v2.1.0).(Clement et al., 2019) For all experimental and bioinformatic steps, a positive control with known editing outcomes, a negative control with no editing and a no template control were processed in parallel with each set of samples.

As has been reported previously, we observed recombination events where double-stranded breaks at the HBB locus were repaired with HBD, a close and nearby homolog of HBB [REF]. These various recombination events could be recognized by the presence of up to 6 SNPs only present in HBD and not the GB101 repair template nor the HBB wildtype sequence.

To estimate the frequency of HBB break repair using HBD as a template we included the fully recombined HBD amplicon sequence (containing all 6 mismatch SNPs relative to the HBB amplicon) as an amplicon in Crispresso (in addition to wildtype HBB and the intended repair outcome with GB101, using the “-a” parameter).

We also wished to quantify both complete recombination (ie amplicons containing all 6 HBD SNPs) and partial recombination (containing 5 or less SNPs). As expected from known HDR pathways, we observed in individual reads that SNPs farther from the cut site were always accompanied by SNPs closer to the cutsite. For Crispresso analysis we added two amplicon sequences consisting of 5 of the 6 HBD-specific SNPs in the 3’ direction and 5 of the 6 SNPs in the 5’ direction from the cutsite. We reasoned that most reads displaying partial HBD recombination should have higher alignment scores to these partially recombined sequences than either the full HBB or full HBD recombination sequences. This strategy was successful in binning reads with regions from HBD but reads with small HBD recombinations at the cutsite

were occasionally binned with WT HBB. These small recombinations could be recognized by the conversion of the cutsite sequence from AC to AA. Therefore, to quantify HBD recombinations we summed the number of reads that mapped to either the full or partial HBD recombinations plus reads containing a mismatch repair at the cutsite without any additional indel. We also added a partially recombined GB101 sequence though we found that very few reads mapped to this suggesting full recombinations are much more likely.

The full list of parameters passed to Crispresso was the same for all analyses:

```
--amplicon_seq \
TCACTAGCAACCTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCC
CTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTA
CAAGACAGG,\
TCACTAGCAACCTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAAAAATCCGCAGTCACTGCC
CTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTA
CAAGACAGG,\
TCACTAGCAACCTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAATCCGCAGTCACTGCC
CTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTA
CAAGACAGG,\
TCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAGACTGCTGTCAATGCC
CTGTGGGGCAAAGTGAACGTGGATGCAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTA
CAAGACAGG,\
TCACTAGCAACCTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGACTGCTGTCAATGCC
CTGTGGGGCAAAGTGAACGTGGATGCAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTA
CAAGACAGG,\
TCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAGACTGCTGTCAATGCC
CTGTGGGGCAAAGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTA
CAAGACAGG \
--amplicon_name \
HBB_WT,\
GPH101,\
GPH101_maxR_5,\
HBD_WT,\
HBD_WT_maxR_5,\
HBD_WT_maxR_3 \
-g CTTGCCCCACAGGGCAGTAA \
--guide_name sgRNA_HBB \
--plot_window_size 50 \
-p 72 \
--skip_failed
```

Summary data for each sample was reported as % WT (unedited), % HDR (incorporation of HBB-SNP donor template), % HBD, %MMEJ (edits that get significantly reduced by POLQ knockdown, as described below), while the rest are classified as “NHEJ”. When presented as % edited alleles, edits are calculated as % of any given edit/(100 – WT).

DNA damage response (DDR markers p21 and yH2AX) analysis

For p21 analysis, 1.5×10^5 cells were spun down at 300 g x 5 min, washed once with PBS, and was resuspended in 22.5 µL of RIPA buffer with 2X Halt protease and phosphatase inhibitors (ThermoFisher). Lysates were incubated on ice for 30 min with intermittent vortexing. Lysates were then spun down in a microcentrifuge at 2000 rpm for 5 min; supernatants were then transferred to fresh tubes. Samples were prepared by mixing 5 µL of protein extract with 1.25 µL of freshly prepared 5X fluorescent master mix as instructed by the ProteinSimple Jess protocol. Samples were denatured for 10 min at 95 °C, quickly spun and loaded into a Jess capillary cartridge. Capillaries were probed with anti-p21 (CST) and anti-alpha tubulin (Abcam) and detected by HRP-conjugated secondary antibodies. Data was normalized to the internal alpha-tubulin loading control and then expressed as FC values over control treatments.

For yH2AX analysis, 1.5×10^5 cells were spun down at 300 g x 5 min. Cells were resuspended in 100 µL of diluted Live/Dead Fixable Violet dye (ThermoFisher, 1:1000 in PBS). Cells were incubated for 20 min in the dark at room temperature. After adding 100 µL of PBS to cell suspension, the cells were spun down and washed once more with 200 µL PBS. The final PBS wash was flicked from the plate and cells were lightly vortexed to resuspend in residual PBS buffer remaining after wash. 100 µL of freshly-prepared 70% ethanol was added to the cell pellets and the plate was tightly sealed with foil, vortexed and then allowed to fix at –20 °C from 1 h to 3 d. Following fixation, 100 µL of cell staining buffer (CSB, BioLegend) was added to cell suspensions and the resuspended cells were spun down at 500 g x 5 min. Cells were washed once more with 200 µL of CSB and then resuspended in 50 µL of CSB and blocked for 15 min at RT. Diluted anti-yH2AX-PE (1:20 in CSB, BioLegend) was added to cells and incubated further for 30 min at RT. Following staining, cells were washed twice with CSB and were immediately analyzed using a Beckman Coulter CytoFLEX flow cytometer.

Measuring editing of OT-1 (NGS-based analysis)

Editing outcomes at the off-target (OT) editing site OT-1 were assessed using an assay very similar to the one described above for measuring the targeted integration of HBB-SNP. The off-target editing site was identified via three different methods (in silico prediction[33], Circle-Seq[43], and Guide-Seq[44]) and was confirmed to be off-target editing site of significance via amplicon sequencing. The workflow for the OT-1 amplicon sequencing assay is very similar to the on-target sequencing assay but involves one less PCR step. A small 166 base pair sequence encompassing the OT-1 editing site is amplified directly from genomic DNA in the first PCR and then tagged with barcode and adapter sequences in a second PCR. The resulting PCR products are sized, quantified, and prepared for sequencing in the same manner as for the HBB on-target sequencing assay. The sequencing data is also processed using the Crispresso2 pipeline[45], but in contrast to the on-target sequencing assay no homology direct repair outcomes are present at OT-1, only % WT and % INDELs are reported.

POLQ knockdown and determination of MMEJ edits

The construct for the shRNA knockdown of PolQ was adapted from the previously reported pLKO[46] and cloned from a third-generation lentiviral plasmid (Lenti SFFV) purchased from Twist Biosciences. An empty vector was constructed to include a DNA stuffer flanked by AgeI and EcoRI restriction enzyme cut sites downstream of a U6 promoter and upstream of a SFFV-EGFP-WPRE cassette (to enable fluorescence-based monitoring of cells expressing the shRNA). shRNA sequences were cloned as duplexed DNA oligos (IDT) into the digested empty vector at the AgeI/EcoRI cut sites using standard Ligation protocols. The target sequences used for POLQ (gene ID: 10721) and non-targeting control (NTC) shRNA were identified using the Broad Genetic Perturbation Platform (GPP) Web Portal (https://portals.broadinstitute.org/gpp/public/gene/search) and are as follows were : POLQ: 5'-GCTGACCAAGATTTGCTATAT-3' and NTC: 5'-CCTAAGGTTAAGTCGCCCTCG-3'.

>MT_vector (GFP)

```
tggaacagatttggaatcacacgacctggatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaacaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggtatataaaa
ttattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacaggcccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgatta
gtgaacggatctcgacggtatcggttaacctcgagcggccgcccccttcaccgagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagt
tttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccggtgtgttgtaaatgagcacacaaaatacacatgctaaaatattatattctatgacctttataaaatcaaccaaaatcttctttttaataactttagtatcaataatta
gaatttttatgttccttttgcaaacttttaataaaaatgagcaaaataaaaaaacgctagttttagtaactcgcgttgttttcttcacctttaataatagctactccaccacttgttcctaagcggtcagctcctgcttcaatcatttttgagcatcttcaaatgttctaactccaccagctgctttaactaaagcattgtctttaaca
actgacttcattagtttaacatcttcaaatgttgcacctgattttgaaaatcctgttgatgttttaacaaattctaatccagcttcaacagctatttcacaagctttcatgatttcttcttttgttaataaacaattttccataatacatttaacaacatgtgatccagctgctttttttacagctttcatgtcttctaaaactaattcataattttt
gtcttttaatgcaccaatatttaataccatatcaatttctgttgcaccatctttaattgcttcagaaacttcgaatgcttttgtagctgttgtgcatgcacctagaggaaaacctacaacatttgttattcctacatttgtgccttttaataattctttacaatagcttgttcaatatgaattaacacaaactgttgcaaaatcaaattca
attgcttcatcacataattgtttaatttcagctttcgtagcatcttgttttaataatgtgtgatctatatatttgtttagtttcatttttctcctatatattcattttaattttaattctttaataatttcgtctactttaactttagcgttttgaacagattcaccaacacctataaaataaatttttagtttaggttcagttccacttgggcgaacagc
aaatcatgacttatcttctaaataaaattttagtaagtcttgtcctggcatattatacattccatcgatgtagtcttcaacattaacaactttaagtccagcaatttgagttaagggtgttgctctcaatgatttcattaatggttcaatttttaatttcttttcttctggtttaaaattcaagtttaaagtgaaagtgtaatatgcacccattt
ctttaaataaatcttctaaatagtctactaatgttttattttgtttttttataaaatcaagcagcctctgctattaatatagaagcttgtattccatctttatctctagctgagtcatcaattacatatccataactttcttcataagcaaaaacaaaatttaatccgttatcttcttctttagcaatttctctacccattcatttaaatccagttaa
agtttttacaatattaactccatatttttcatgagcgattctatcacccaaatcacttgttacaaaacttgaatatagagccggatttttttggaatgctatttaagcgttttagatttgataattttcaatcaattaaaattggtcctgtttgatttccatctaatcttacaaaatgaccatcatgttttattgccattccaaatctgtcagca
tctgggtcattcataataataatatctgcatcatgtttaataccatattcaagcggtatttttcatgcaggatcaaattctggatttggatttacaacatttttaaatgtttcatcttcaaatgcatgctcttcaacctcaataacgttatatcctgattcacgtaatatttttggggtaaatttagttcctgttccattaactgcgctaaaa
ataatttttaaatctttttagcttcttgctctttttgtagaattctcgacctcgagacaaatggcagtattcatccacaattttaaaagaaaaggggggattgggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaatttt
ggctcccgatcgttgcgttacacacacaattactgctgatcgagtgtagccttcgaatgaaagaccccacctgtaggtttggcaagatagctgcagtaacgccatttgcaaggcatggaaaaataccaaaccaagaatagagaagttcagatcaagggcgggtacatgaaaatagctaacgttgggccaaa
caggatatctgcggtgagcagtttcggccccggcccggggccaagaacagatggtcaccgcagtttcggccccggcccgaggccaagaacagatggtccccagatatggcccaaccctcagcagtttcttaagacccatcagatgtttccaggctcccccaaggacctgaaatgaccctgcgccttatttga
attaaccaatcagcctgcttctcgcttctgttcgcgcgcttctgcttcccgagctctataaaagagctcacaacccctcactcggcgcgccagtcctccgattgactgagtcgccctgatcattgtcgatcctaccatccactcgacacacccgccaggatccccagcggccgcatggtgagcaagggcgaggagc
tgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtg
cttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggagg
acggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctg
ctgcccgacaaccactacctgagcacccagtccaagctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaagctagcgactatcatatgcttaccgtaatcaacctctggattacaaaatttgtgaaa
gattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttg
gggcattgccaccacctgtcagctcctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgc
gggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttgggccgcctccccgcactagtgtctcgagggcccgcatcgagctcggtacctttaagaccaatgacttacaa
ggcagctgtagatcttagccactttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatctgctttttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagta
gtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagtgtggaaaatctctagcagtcctggccaacgtgagcaccgtgctgacctccaaatatcgttaagctggagcctgggagccggcctggccctccgcccccccacccccgcagcccacccctggtctttgaataaag
tctgagtgagtggccgacagtgcccgtggagttctcgtgacctgaggtgcagggccggcgctagggacacgtccgtgcacgtgccgaggcccccctgtgcagctgcaaggacaggcctagccctgcaggcctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgcctcatggctg
actaattttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggacgcttttttggaggccgaggcttttgcaaagatcgaacaagagacaggacctgcaggttaattaaatttaaatcatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttg
ctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgc
tttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtat
gtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaag
gatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggca
```

cctatctcagcgatctgtctatttcgttcatccatagttgcatttaaatggccggcctggcgccgctttaaacctagatattgatagtctgatcggtcaacgtataatcgagtcctagctttgcaaacatctatcaagagacaggatcagcaggaggctttcgcatgagtattcaacatttccgtgtcgcccttattccctttt
ttgcggcattttgccttcctgttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcgcgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgctttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatccc
gtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtattcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaagga
gctaaccgctttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgtaaactattaactggcgaactacttactctagcttcccggcaacaattgatagactggatgga
ggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagat
cgctgagataggtgcctcactgattaagcattggtaaccgattctaggtgcattggcgcagaaaaaaaatgcctgatgcgacgctgcgcgtcttatactcccacatatgccagattcagcaacggatacggcttccccaacttgcccacttccatacgtgtcctcctaccagaaatttatccttaagatcccgaatcgtt
taaacgcgatcgacgtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggagtattta
cggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgt
ggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggtctctctgg
ttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagtgtggaaaatctctagcagtcctggccaacgtgagcaccgtgctgacctccaaatatcgttaagctggagcctgggagccggcctggccctccgcccc
CD34+ HSPC cells were transduced using lentivirus produced using shRNA transfer vectors at day 1 post thaw using methods described above and MOIs of 2.5 - 7.5. Three days after transduction, cells were edited in duplicate at HBB as described above, using HBB-SNP or HBB-UbC-mCherry donor AAV6 (MOIs of 75 - 2.5×10^4 vector genomes/cell). Three days post editing, cells were pelleted and resuspended in media with DAPI (Miltenyi Biotec).

For cells edited with HBB-UbC-mCherry: rates of integration of the donor were measured using a Beckman Coulter CytoFLEX. DAPI (Miltenyi Biotec) was used to discriminate live and dead cells. GFP expression was used to differentiate transduced cells from untransduced cells and rates of mCherry integration were compared between the two populations to quantify the impact of lentiviral-based shRNA expression on HDR rates.

For cells edited with HBB-SNP: single, live, GFP+ (shRNA+) and GFP- (shRNA-, negative control) cells were collected using a FACSAria cell sorter (Becton Dickinson); purity of populations was confirmed by post-sort purity checks. Post sort, genomic DNA was harvested from each sorted cell population and editing outcomes were determined using the pipeline outlined above. To determine which edits were reduced by POLQ knockdown, a one-sided t-test comparing the GFP+ and GFP- conditions for each individual editing outcome. Those edits with FDR-corrected (Benjamini-Hochberg) p-value below 0.1 were labeled as POLQ-dependent on “MMEJ”.

LT-HSC sorting Method and Materials

***Table M1*: Antibodies and reagents for LT-HSC sort**

Fluorophore Conjugate/Dye	Antigen/Target Specificity	Clone	Vendor

Alexa 488	CD34	581	Biolegend
PE-Cy7	CD38	HIT2	BD Biosciences
BV510	CD45RA	HI100	Biolegend
PE	CD49c	ASC-1	Biolegend
BV421	CD49f	GoH3	Biolegend
BV711	CD90	5E10	Biolegend
APC	CD201	RCR-401	Biolegend
Alexa 700	CD45	HI30	Biolegend
DyLight 800 Maleimide	Live/Dead	N/A	Thermo Scientific

Cryopreserved samples were rapidly thawed in warm GMP SCGM (CellGenix) media, washed with cell staining buffer (Biolegend). Washed cells were incubated with a panel of fluorochrome-conjugated anti-human monoclonal antibodies (mAb) and viability dye to characterize hematopoietic stem cell compartments. The following directly conjugated mAbs used in this study were obtained from BD Biosciences: CD38-PE-Cy7 (Clone HIT2), Biolegend: CD34-Aexa488 (581), CD45RA-BV510 (HI100), CD49c-PE (ASC-1), CD49f-BV421 (GoH3), CD90-BV711 (5E10), CD201-APC (RCR-401), CD45-Alexa700 (HI30) and Thermo Fisher Scientific: DyLight 800 Maleimide. Brilliant Stain Buffer Plus (BD Bioscience) was added to stabilize the fluorophore-conjugated antibody cocktail. Cells were stained for 30 min at 4 °C and washed with cell staining buffer and acquired within 1 h on a custom SORP five laser FACSAria Fusion (BD Biosciences).

FACSAria Fusion was calibrated with Cytometer Setup and Tracking beads (BD Biosciences, 655050), the sort parameters were set to 20psi with a 100 µm nozzle, and the droplet stream was calibrated with Accudrop Beads (BD Biosciences, 345249). Sort layout was set to 4-way purity and four subfractions were collected into 5 mL FACS tubes as follows:

1) Long Term Hematopoietic Stem Cell (LT-HSC) enriched (CD34+CD45RA-CD90+CD201+CD49f+CD49c+)
2) Short Term hematopoietic Stem Cell (ST-HSC) enriched (CD34+CD45RA-CD90+CD201- (CD49f-CD49cdim))
3) Hematopoietic Stem and progenitor cell (HSPC) enriched (CD34+CD45RA-CD90-CD201- (CD49f-CD49c-)
4) Lineage committed progenitors (CD34+CD45RA+ (CD90dimCD201-CD49f-CD49cdim)

Aliquots of sorted sample populations were re-acquired to assess the purity of the sort. Sorted cells were spun down, supernatant aspirated, and snap frozen at -80 °C for DNA extraction and NGS analysis.
The bulk cells were phenotyped for cell sorting with the surface markers CD45, CD34, CD45RA, CD201, CD90, CD49f, CD49c[9,35–37]. A physical gate was applied to remove debris and isolate HSC sized cells, doublet cells were removed with a SSC-singlet and FSC-singlet gates, dead cells were removed, the CD45+ cells were sub-fractionated into HSC/HSPC (CD45+CD34+CD45RA-) and linage committed (sort population 4: CD45+CD34+CD45RA+) compartments. The HSC/HSPC was further divided into HSPC (sort population 3: CD45+CD34+CD45RA-CD90-CD201-(CD49f-CD49c-)), short-term HSC (sort population 2: CD45+CD34+CD45RA-CD90+CD201-(CD49f-CD49cdim)) and long-term HSC (sort population 1: CD45+CD34+CD45RA-CD90+CD201+CD49f+CD49c+).

CFU progenitor assay

At 48 h post gene editing, 250 cells per well were plated in Methocult Optimum media in SmartDish plates (both StemCell Technologies). Plates were incubated in a secondary enclosure at 37 °C, 5% CO_2, and 5% O_2 for 14 d before scoring colonies using the human mPB program on a STEMvision imager (StemCell Technologies).

Measuring targeted integration of HBB-SNP in Colonies (CFU-seq)

Individual colonies were picked and gDNA was extracted using Lucigen Quickextract kit according to manufacturer's instructions. NGS library prep on gDNA was performed as described in above section titled Measuring Targeted Integration of HBB-SNP.

Raw fastq files output from the sequencer were analyzed using our in-house On-target HBB CFU Bioinformatics Pipeline. This pipeline uses Crispresso 2 (v2.1.0) to quantify the various gene editing outcomes in each colony. Parameters for Crispresso 2 were set to be identical to those used for On-Target CD34 NGS analysis. Output counts and fractions of each allele from Crispresso analysis was used to infer genotypes. Filters were applied to remove low quality colonies. Colonies with fewer than 2000 reads aligned were removed as the low read count would likely impact quantification. A 10% fraction threshold was used to call the presence of expected alleles. Colonies with more than two alleles above the 10% fraction threshold were removed as these were likely not single clones. NoCall colonies included any colonies that did not produce a band on In/Out PCR or were removed by the above bioinformatics filters.

Guide-seq

Genomic DNA (gDNA) was isolated using PureBind Blood Genomic DNA Isolation Kit (Ocean Nanotech), quantified, library preparation was performed, and quality was assessed (LAB-SOP-018, GeneGoCell) then sequenced (NextSeq2000, Illumina). Raw sequence reads were demultiplexed into sample-specific fastq files (bcl2fastq program v2.20.0.422, Illumina). The resulting fastq files were processed as follows: Low-quality reads were removed using quality score threshold 28 (Q28), and PCR duplicates were removed using the UMIs. The resulting fastq files were analyzed to generate quality control (QC) statistics. Reads were aligned to the human genome (hg38) using BWA v0.7.17- r1188 (GeneGoCell NGS bioinformatics pipeline v2.2.3).

The control and experimental samples were further analyzed using the same process, abbreviated here: For a given site, the dsODN insertion rate was calculated as the number of site-specific reads with dsODN incorporation vs. total number of site-specific reads. The alignment results were analyzed using G-GUIDE analysis program v4.0 to generate the genome-wide dsODN insertion sites and report break points (BPs) for each high-quality read. Control sample background sites were subtracted from the edited samples, and only sample-specific sites are reported.

Karyotyping

Cryopreserved aliquots of 2M cells were used for each submission. Aliquots of 2M cells for each sample (24h post-editing) were prepared by centrifuging cells (180g, 7 min) and resuspending in Cryostor CS10 solution (BioLife Solutions) at a density of 10M cells/mL. Frozen cell aliquots were then sent to KromaTiD (Longmont, CO) for karyotyping (G-banding). Briefly, after harvest and fixation, the fixed cells were washed twice with fixative (prepared fresh day-of-use) and the O.D. was adjusted. Drops of the final cell suspension were placed on clean slides and aged for 60 min at 90 °C. Slides were digested in a pancreatin solution with Isoton II diluent. The enzymatic reaction was then stopped by rinsing with FBS, followed by application of a stain solution (3:1 Wright/Gurr buffer) which was poured on the slides so that it covered the entire surface. After staining for up to 1 min, slides were washed with de-ionized water for 1-5 s and air dried. A mounting medium was applied to the slides and sealed with a coverslip. The slides were scanned on the microscope for cell analysis.

Translocation-seq

Detection of the sequence of interest and their translocated partners, in this case our editing site, known off-target site and their translocation partners. Samples were treated as follows: gDNA was isolated (PureBind Genomic DNA Isolation Kit, Ocean Nanotech), fragmented via sonication, followed by DNA-end repair, UMI adapter ligation, and PCR amplification enrichment of fragments that contain editing targets and translocations, then prepared for sequencing (LAB-SOP-017, GeneGoCell). The amplified gDNA fragment library was sequenced (NextSeq 2000, Illumina) and DNA sequence generation via sequencing-

by-synthesis (SBS) (LAB-SOP-022, GeneGoCell), demultiplexed (bcl2fastq v2.20.0.422, Illumina) and processed as described next (v2.0.9, GeneGoCell):

GeneGoCell's proprietary G-Trans platform was used to amplify and quantify all potential translocations in an unbiased manner. Translocations were quantified between the target listed below to anywhere else across the genome (hg38). Low quality reads and PCR duplicates were removed via Q28 and UMIs, respectively. Quality control was run (v.0.10.1, FastQC), reads were aligned to hg38 (v0.7.17-r1188, BWA), and results were analyzed (proprietary translocation analysis v1.6, GeneGoCell) to identify potential genome-wide translocation sites. Donor and recipient genomic loci BPs were calculated per read. Called BPs met the following CRISPR/Cas9 genome editing associated criteria: >= 10 UMI reads, a minimum of 3 BPs in the flanking 200 bp of the peak, position +/- 100 base pairs, and peak BP:total region read counts ratio <0.9. The on-target translocation rate required an additional manual calculation to remove the DNA donor sequence:

Translocation rate = $(Number\ of\ Reads\ for\ th\ region / Number\ of\ Target\ Specific\ Reads)$ x 100

***Table M2*: Trans-seq target list**

Chr	0-start	End	Name
Chr11	5226967	5226990	HBB
Chr9	101833583	101833606	OT-1

***Table M3*: Trans-seq primer list**

Chromosome	Primer ID	Sequence (5'-3')	Description
Chr11	GRB106-HBB-R-O-1	CATAAAAGTCAGGGCAGAGC	Outer Right
Chr11	GRB106-HBB-R-I-1	GCTTACATTTGCTTCTGACACAACT	Inner Right
Chr11	GRB106-HBB-L-O-1	TGTCTCCACATGCCCAGTT	Outer left
Chr11	GRB106-HBB-L-I-1	GGGCCTCACCACCAACTT	Inner Left
Chr9	GRB106-OT1-L-O-1	CTTTCCCGTTCTCCACCCA	Outer Right
Chr9	GRB106-OT1-L-I-1	GCTATGGAAAGGGGAAGATCC	Inner Right
Chr9	GRB106-OT1-R-O-1	ACGGCCTAAGAAATTATAGTTTAGCA	Outer left
Chr9	GRB106-OT1-R- I-1	CAGTATGTCCAACTCCCAAATTGA	Inner Left

scRNA-seq

To assess gene expression profiles from single cells, 2 million cryopreserved cells were thawed for use with 10X Genomics Chromium Next GEM Single Cell 3' Gene Expression Reagent Kit (10X 3' Kit). The thawed cells were counted with AO/PI viability stain on the Nexcelom cell counter. Approximately, 8000 live cells were added to a master mix for reverse transcription (RT Reagent B, Template Switching Oligo, Reducing Agent B, and RT enzyme C), then loaded into a Chromium Next GEM Chip G for running in the Chromium Controller to generate Gel Beads-in-emulsion (GEMs). The GEMs were transfer to tubes for RT incubation in a Bio-Rad C1000 Touch for 45 min at 53 °C, then 5min at 85 °C and held at 4 °C. After RT, the GEMs were purified with Dynabeads™ MyOne™ SILANE. The eluted cDNA was amplified by using the Amp Mix and cDNA primers with 11 cycles of PCR in a Bio-Rad C1000 Touch. The dsDNA cDNA product was analyzed using the High Sensitivity DNA Chip on an Agilent Bioanalyzer 2100. 10uL of the dsDNA cDNA product was fragmented with the fragmentation primer, end-repaired and A-tailed to prepare for the ligation of the sequencing adapters. Afterwards, the dsDNA was purified with a double-sided SPRI. Illumina sequencing adapters were ligated to the dsDNA to generate the sequencing library. Another 15 cycles of PCR in a Bio-Rad C1000 Touch was used to amplify and index the sequencing library.

The indexed libraries were purified with a double-sided SPRI and qualitatively measured with the High Sensitivity DNA Chip on an Agilent Bioanalyzer 2100 to assess the size and the concentrations were measured by Qubit Broad Range kit. Each sample library was normalized to 900 pM and pooled in equal volumes. The library pool plus including 10% PhiX control library was denatured and then loaded onto a P3 flowcell on the Illumina NextSeq 2000. The run parameters were Read 1: 28, Index 1: 10, Index 2: 10, Read 2: 90, per 10X 3' Kit protocol. After the run, the sequence metrics was checked to see read quality and then bcl files were converted to fastq. The fastq were then input into the Graphite single cell pipeline for analysis.

Data processing and analysis were performed in R version 4.2.0 via RStudio, using Seurat (v4.3.0). Visualizations were created with dittoSeq (v1.8.1) (https://github.com/dtm2451/dittoSeq/) and ggplot2. Seurat's *Read10X* function was used to generate a count data matrix using the filtered count matrix genes and cells, gene names, and barcode files provided by 10X. A Seurat object was created with the count data matrix and metadata and filtered to keep genes present in at least 3 cells and cells meeting cohort selection criteria of at least 200 genes. Log normalization was performed using Seurat's *NormalizeData* function with a scale factor of 10,000, and highly variable features were identified using Seurat's *FindVariableFeatures* The data matrix was then scaled using Seurat's *ScaleData* function with nCount_RNA regressed out, and dimensionality reduction through Uniform Manifold Approximation and Projection (UMAP) was performed with the appropriate dimensions selected based on the corresponding principal component analysis (PCA) elbow plot.
The Seurat funciton *RunAzimuth* was used as reference-based mapping to annotate the data to the Human bone marrow reference (https://azimuth.hubmapconsortium.org/)

Isolation of CD34+ cells

Leuokopaks were purchased from AllCells LLC and these were collected from healthy donors per standard protocols using mobilization with G-CSF+ Plerixafor. CD34+ cells were isolated from the leukopaks within 24 h by first removing the platelets using the LOVO cell Processing System (Fresenius Kabi). GMP grade reagents, buffers and columns for CD34 immunomagnetic selection were purchased from Milteny Biotec and the platelet washed cells were incubated for 30-35 min using the CD34 Reagent following which a subsequent wash for excess antibodies was performed on the LOVO. The washed and labelled cells were subject to immunomagnetic selection using the CliniMACS Plus instrument (Miltenyi Biotec) following which the cells were cryopreserved at a concentration of $5x10^6$-$1x10^7$ cells/mL in CryoMACS 50 or 250 Bags (Miltenyi Biotec) for the gene edited drug product generation step.

Large scale editing of HSPCs

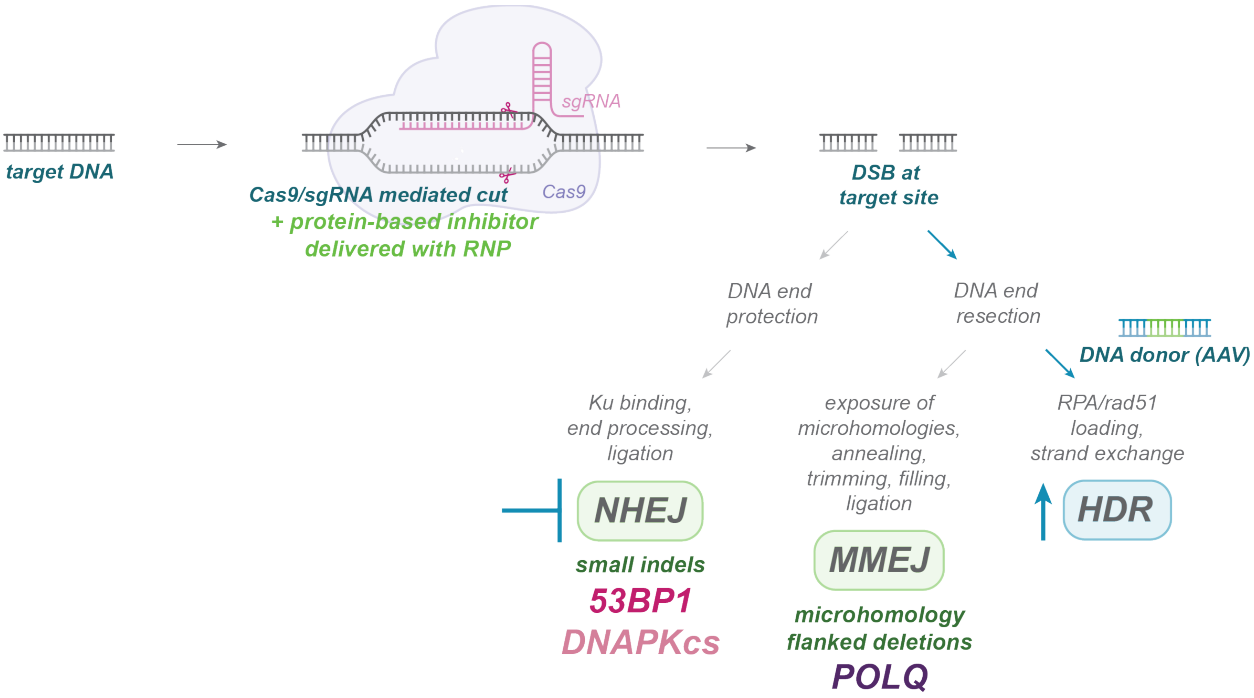
At least $5x10^7$-$3x10^8$ Cryopreserved CD34+ HSPCs were then thawed at 37 °C and cultured in supplemented cytokine rich SCGM media (CellGenix) containing recombinant cytokines at 100 ng/mL each Flt-3L, TPO, and SCF (PeproTech) and 35nM UM171 (ExCellThera) in gas permeable vessels and incubated in 5% CO_2+ 5% O_2 for 48-72 h. The cells were then washed and resuspended in 3-10 mL of electroporation buffer (Hyclone). A GMP grade chemically modified single guide RNA (sgRNA) targeting the HBB locus was purchased from Agilent with modifications for 2'-O-methyl-3'-phosphorothioate at the three terminal nucleotides of the 5' and 3' ends with the sequence 5'-CTTGCCCCACAGGGCAGTAA-3'. The gene editing reagents were pre-complexed as an RNP containing 2 mg/mL sgRNA (Agilent Technologies) and 10mg/mL SpyFi Cas9 (Aldevron) at a 2.5:1 molar ratio for 10 minutes at room temperature. Approximately 169 µL of RNP was added per 1 mL of cell suspension in electroporation buffer. For conditions testing the HDR booster, thawed i53 variant protein was mixed well by pipetting and added to the RNP at a concentration of 0.8 mg/mL (of total electroporation volume) following which the cells were electroporated using the MaxCyte GTx system using the CL1.1 or CL2 closed cartridge that are suitable for GMP manufacturing. Following electroporation, the cells were allowed to rest for 10 minutes in an incubator at 37 °C. In the meanwhile, prepared AAV6.133K virus carrying the corrected sequence for HBB was thawed and added at either $6.25x10^2$ or $1.25x10^3$ vector genomes/cell into culture media following which the electroporated cells were split equally (for different MOI conditions) and transferred to gas permeable culture vessels. At 16-24 h post-gene editing, the cells from each condition were collected and centrifuged at 300xg for 10 min to pellet the cells. The supernatant was aspirated, and the cell pellet(s) were washed with and re-suspended in PlasmaLyte buffer with 2% (v/v) HSA. A cell count was performed using the

NC202 counter that uses AO/DAPI staining using the pre-set Cell count and Viability protocol. Cell counts were used to determine cell yield, viability and concentrations for cryopreservation. Following a final centrifugation step at 450xg for 10 min, the cells were resuspended in cold cryopreservation media CryoStor CS5 (BioLife Technologies) and aliquoted into vials at a final concentration of $5x10^6$-$1.2x10^7$ cells/mL. The vials were then subject to cryopreservation using a controlled rate freezer and storage in vapor phase LN_2 at ≤-150 °C prior to performing all analytical metrics.

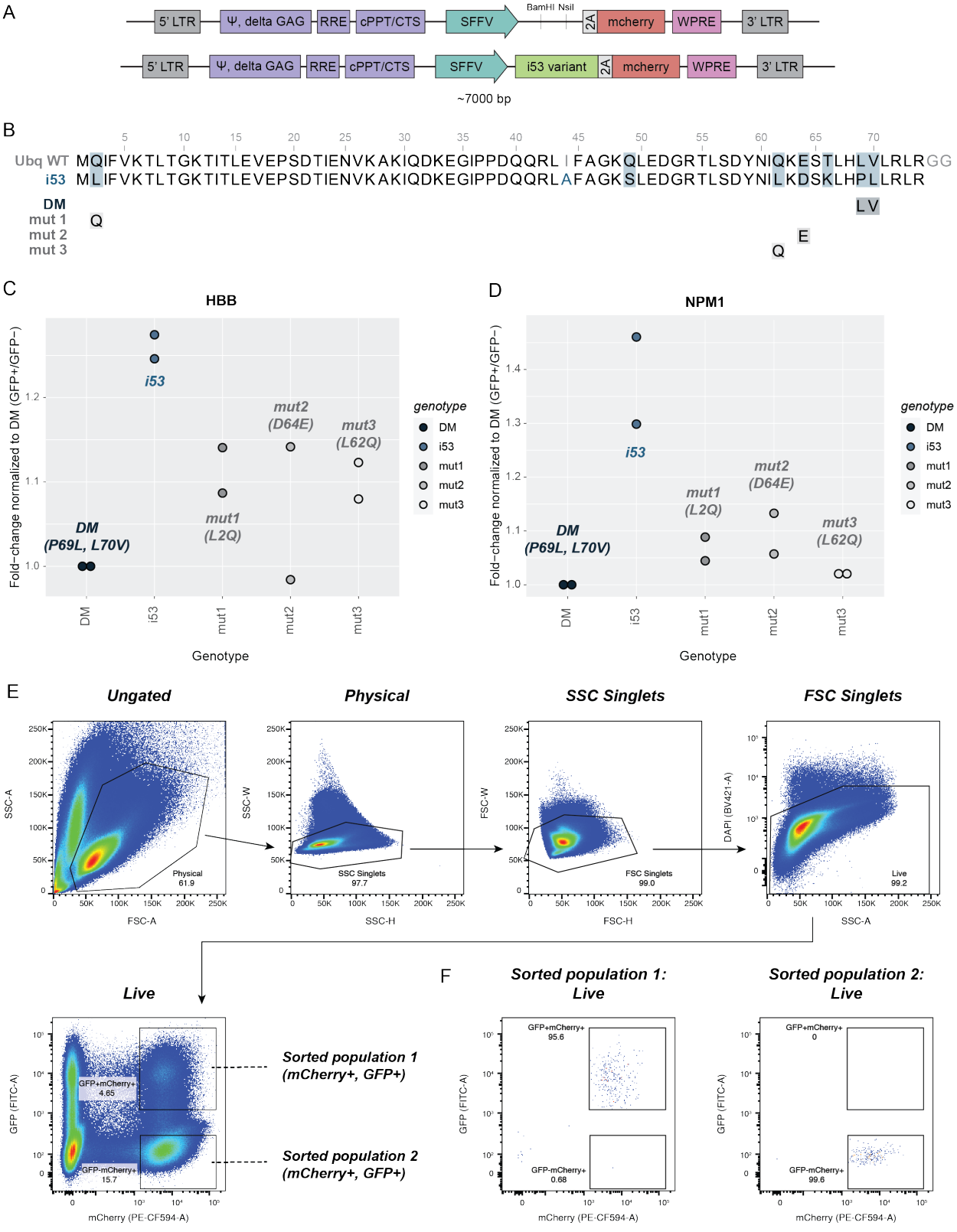
References

1. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578,** 229–236 (2020).
2. Porteus, M. H. A New Class of Medicines through DNA Editing. *N. Engl. J. Med.* **380,** 947–959 (2019).
3. Morgan, R. A., Gray, D., Lomova, A. & Kohn, D. B. Hematopoietic stem cell gene therapy: progress and lessons learned. *Cell Stem Cell* **21,** 574–590 (2017).
4. Kohn, L. A. & Kohn, D. B. Gene therapies for primary immune deficiencies. *Front. Immunol.* **12,** 648951 (2021).
5. Ferrari, G., Thrasher, A. J. & Aiuti, A. Gene therapy using haematopoietic stem and progenitor cells. *Nat. Rev. Genet.* **22,** 216–234 (2021).
6. Nambiar, T. S., Baudrier, L., Billon, P. & Ciccia, A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell* **82,** 348–388 (2022).
7. Ferrari, S. *et al.* Gene editing of hematopoietic stem cells: hopes and hurdles toward clinical translation. *Front. Genome Ed.* **3,** 618378 (2021).
8. Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21,** 1468–1478 (2019).
9. Shin, J. J. *et al.* Controlled Cycling and Quiescence Enables Efficient HDR in Engraftment-Enriched Adult Hematopoietic Stem and Progenitor Cells. *Cell Rep.* **32,** 108093 (2020).
10. Ferrari, S. *et al.* 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 (2022).
11. Allen, D. *et al.* High-Throughput Imaging of CRISPR- and Recombinant Adeno-Associated Virus-Induced DNA Damage Response in Human Hematopoietic Stem and Progenitor Cells. *The CRISPR Journal* **5,** 80–94 (2022).
12. Shy, B. R. *et al.* High-yield genome engineering in primary cells using a hybrid ssDNA repair template and small-molecule cocktails. *Nat. Biotechnol.* **41,** 521–531 (2023).
13. De Ravin, S. S. *et al.* Enhanced homology-directed repair for highly efficient gene editing in hematopoietic stem/progenitor cells. *Blood* **137,** 2598–2608 (2021).
14. Riesenberg, S. & Maricic, T. Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. *Nat. Commun.* **9,** 2164 (2018).
15. Canny, M. D. *et al.* Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36,** 95–102 (2018).
16. Dever, D. P. *et al.* CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature* **539,** 384–389 (2016).
17. Wilkinson, A. C. *et al.* Cas9-AAV6 gene correction of beta-globin in autologous HSCs improves sickle cell disease erythropoiesis in mice. *Nat. Commun.* **12,** 686 (2021).
18. Hoban, M. D. *et al.* Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood* **125,** 2597–2604 (2015).
19. Cromer, M. K. *et al.* Gene replacement of α-globin with β-globin restores hemoglobin balance in β-thalassemia-derived hematopoietic stem and progenitor cells. *Nat. Med.* **27,** 677–687 (2021).
20. Gomez-Ospina, N. *et al.* Human genome-edited hematopoietic stem cells phenotypically correct Mucopolysaccharidosis type I. *Nat. Commun.* **10,** 4045 (2019).
21. Pavel-Dinu, M. *et al.* Gene correction for SCID-X1 in long-term hematopoietic stem cells. *Nat. Commun.* **10,** 1634 (2019).
22. Mateos-Gomez, P. A. *et al.* Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518,** 254–257 (2015).
23. Ramsden, D. A., Carvajal-Garcia, J. & Gupta, G. P. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nat. Rev. Mol. Cell Biol.* **23,** 125–140 (2022).
24. Xu, L. *et al.* Molecular Dynamics of Genome Editing with CRISPR/Cas9 and rAAV6 Virus in Human HSPCs to Treat Sickle Cell Disease. *Molecular Therapy - Methods & Clinical Development* (2023). doi:10.1016/j.omtm.2023.07.009
25. Collins, P. L. *et al.* DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. *Nat. Commun.* **11,** 3158 (2020).
26. Tatiossian, K. J. *et al.* Rational Selection of CRISPR-Cas9 Guide RNAs for Homology-Directed Genome Editing. *Mol. Ther.* **29,** 1057–1069 (2021).
27. Robert, F., Barbeau, M., Éthier, S., Dostie, J. & Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **7,** 93 (2015).
28. Yue, X., Bai, C., Xie, D., Ma, T. & Zhou, P.-K. DNA-PKcs: A Multi-Faceted Player in DNA Damage Response. *Front. Genet.* **11,** 607428 (2020).
29. Zenke, F. T. *et al.* Pharmacologic Inhibitor of DNA-PK, M3814, Potentiates Radiotherapy and Regresses Human Tumors in Mouse Models. *Mol. Cancer Ther.* **19,** 1091–1101 (2020).
30. Berger, M. *et al.* BAY-8400: A Novel Potent and Selective DNA-PK Inhibitor which Shows Synergistic Efficacy in Combination with Targeted Alpha Therapies. *J. Med. Chem.* **64,** 12723–12737 (2021).
31. Fok, J. H. L. *et al.* AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. *Nat. Commun.* **10,** 5065 (2019).

32. Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **24,** 1216–1224 (2018).
33. Cradick, T. J., Qiu, P., Lee, C. M., Fine, E. J. & Bao, G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol. Ther. Nucleic Acids* **3,** e214 (2014).
34. Lattanzi, A. *et al.* Development of β-globin gene correction in human hematopoietic stem cells as a potential durable treatment for sickle cell disease. *Sci. Transl. Med.* **13,** (2021).
35. Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1,** 635–645 (2007).
36. Tomellini, E. *et al.* Integrin-α3 Is a Functional Marker of Ex Vivo Expanded Human Long-Term Hematopoietic Stem Cells. *Cell Rep.* **28,** 1063–1073.e5 (2019).
37. Fares, I. *et al.* EPCR expression marks UM171-expanded CD34+ cord blood stem cells. *Blood* **129,** 3344–3351 (2017).
38. Ferreira da Silva, J., Meyenberg, M. & Loizou, J. I. Tissue specificity of DNA repair: the CRISPR compass. *Trends Genet.* **37,** 958–962 (2021).
39. Fu, Y.-W. *et al.* Dynamics and competition of CRISPR-Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Res.* **49,** 969–985 (2021).
40. Hussmann, J. A. *et al.* Mapping the genetic landscape of DNA double-strand break repair. *Cell* **184,** 5653–5669.e25 (2021).
41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **60,** 2126–2132 (2004).
42. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **53,** 240–255 (1997).
43. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14,** 607–614 (2017).
44. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–197 (2015).
45. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37,** 224–226 (2019).
46. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124,** 1283–1298 (2006).

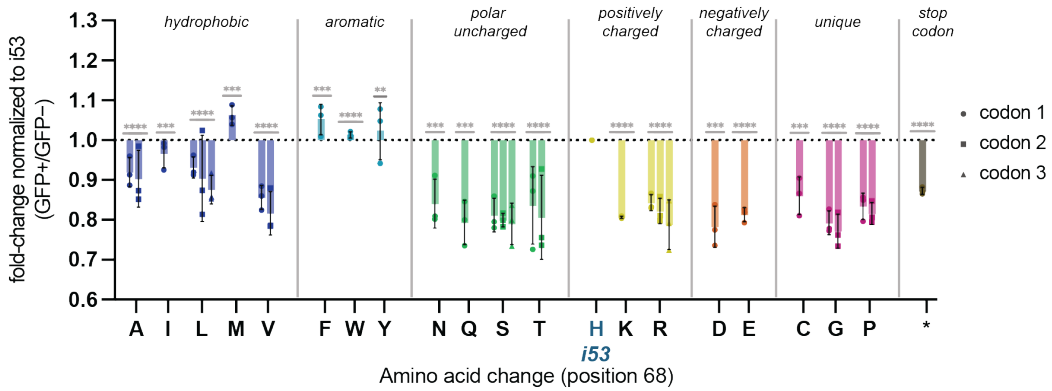


***Supplemental Figure 1.1*: Boosting HDR-based gene editing outcomes through protein-based inhibitors of key DNA repair enzymes.** Schematic outlining the impact antagonists of key DNA repair enzymes can have on the various editing outcomes that occur after a Cas9-mediated site-specific double strand break (DSB). Key target enzymes for NHEJ and MMEJ pathways are listed below. Inhibition of 53BP1 or DNAPKcs is predicted to increase HDR, as shown in blue arrows.

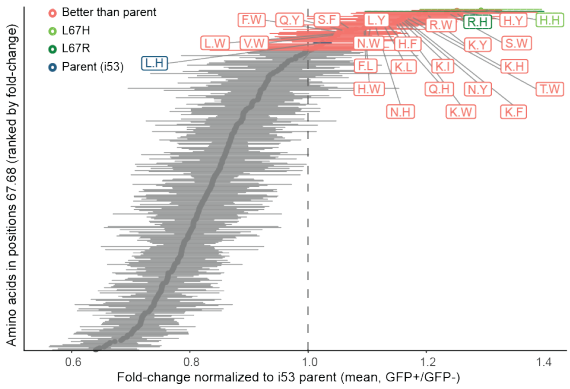


***Supplemental Figure 1.2*: Lentiviral pooled screening design and validation.** (A) A lentiviral transfer plasmid was built with restriction sites upstream of a T2A-mCherry-WPRE cassette to easily clone in sequences of interest. Placement of a T2A-mCherry tag at the 3' end of the protein variants allowed for independent expression of both the protein variant and mCherry post translation, enabling fluorescence-based monitoring of cells expressing protein variants of interest. (B) To validate the lentiviral-based screening system in HSPCs, different variants of i53 were cloned into the lentiviral vector: i53 (positive control), a previously reported dead variant of i53 (DM, negative control), and three variants of i53 that have been previously reported to have decreased (but detectable) binding to the 53BP1 Tudor domain relative to i53 ("mut1" = L2Q, "mut2" = D64E, and "mut3" = L62Q). Plasmids encoding these five i53 variants were pooled together to generate a mock library. CD34+ HSPC cells were transduced using lentivirus packaged with the mock library and edited in duplicate at the *HBB* locus using HBB-UbC-GFP AAV6 (MOI = 2500) and *NPM1* locus using NPM1-GFP AAV6 (MOI = 2500). NGS analysis of the gDNA purified from sorted mCherry+GFP+ and mCherry+GFP- populations indicated differential variant enrichment in the populations relative to the DM control for editing at (C) *HBB* and (D) *NPM1*. (E) Representative flow cytometry plots showing the gating strategy used to sort edited GFP positive cells and GFP negative cells in transduced (mCherry+) HSPC populations. (F) Purity of sorted populations was confirmed by post-sort purity checks.

A

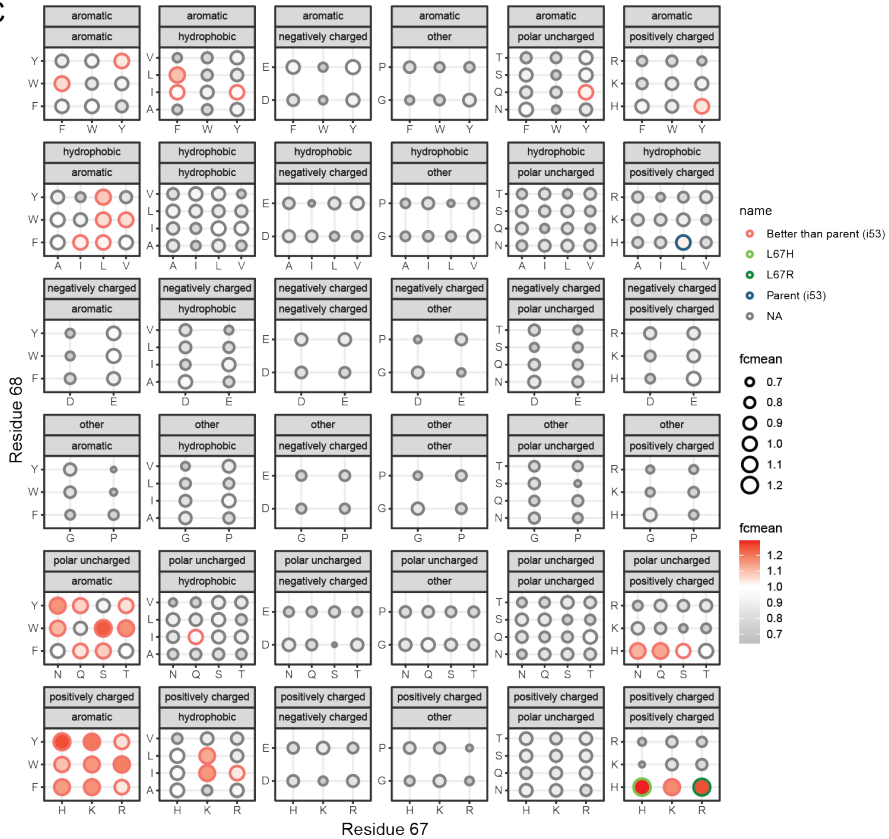


B



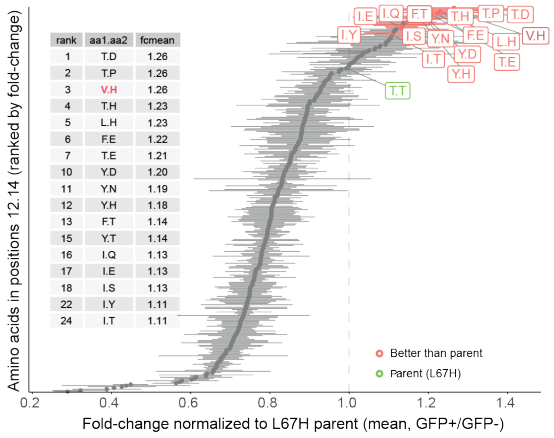
rank	aa1.aa2	fcmean
1	H.H	1.29
2	H.Y	1.25
3	R.H	1.24
4	S.W	1.23
5	K.Y	1.20
6	R.W	1.19
7	K.H	1.17
8	T.W	1.16
9	K.F	1.16
10	N.Y	1.16
11	K.I	1.15
12	K.W	1.15
13	H.F	1.15
14	Q.H	1.13
15	K.L	1.13
16	N.H	1.11
17	N.W	1.11
18	H.W	1.10
19	F.L	1.09
20	L.Y	1.08
21	S.F	1.07
22	Q.Y	1.07
23	F.W	1.05
24	L.W	1.05
25	V.W	1.05

C



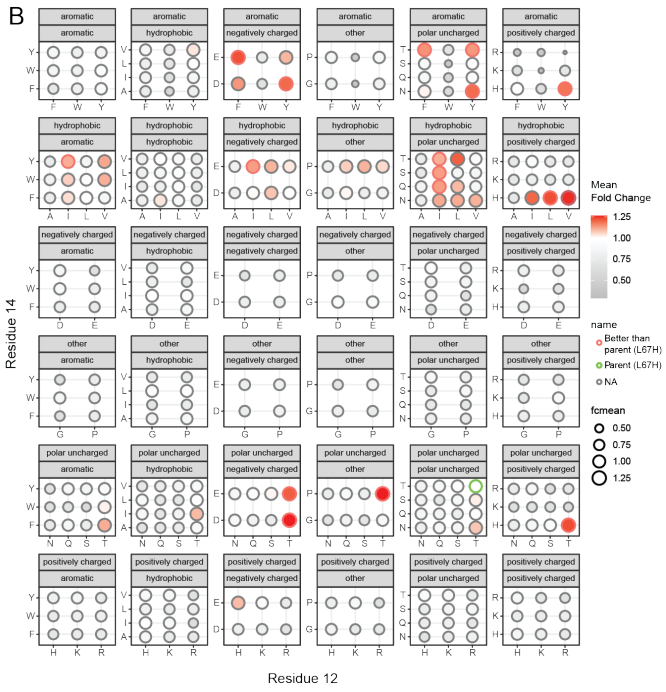
Supplemental Figure 1.3 **(related to Figure 1C): Screening libraries targeting i53 residues 67 and 68**. (A) A focused saturation mutagenesis library was generated using NNK primers to independently vary the amino acid identity of residue 68 of i53. Differential enrichment of variants containing mutations at residue 68 in the mCherry+GFP+ population was calculated from the variant abundance in sorted mCherry+GFP+ and mCherry+GFP- populations relative to i53 (H68). CD34+ HSPC cells were transduced using lentivirus packaged with this focused NNK library and edited at the HBB locus using HBB-UbC-GFP AAV6 (MOI = 1250). *N = 3* separate pooled analyses. Each bar represents a unique codon for that amino acid. ***: p-val < 0.01; ***: p-val < 0.001' ****: p-val < 0.0001. T*wo-tailed t-test with Holm-Šídák correction for multiple comparisons. (B) A combinatorial library was designed to explore all amino acid combinations (excluding cysteine and methionine, N = 324) at i53 parent sequence L67 and H68 at the 53BP1/i53 binding interface. CD34+ HSPC cells were transduced using lentivirus packaged with the combinatorial library and edited in triplicate at the *HBB* locus using HBB-UbC-GFP AAV6 (MOI = 2500). Differential variant enrichment was calculated from the variant abundance in sorted mCherry+GFP+ over mCherry+GFP- and ranked relative to parent i53, shown in blue. Hits L67R and L67H from the NNK screen are shown in green, respectively. Variants withaverage fold change over parent (i53) larger than 1.0 are highlighted in red. (C) Dot plot representation of variant fold change enrichment as in (B), clustered by amino acid properties (amino acid variations of residues 67 and 68 shown on the x-axis and y-axis, respectively). Note amino acid properties listed for each cluster shown refer to those for residue 67 above and to residue 68 below. *N = 3* separate pooled libraries.

A

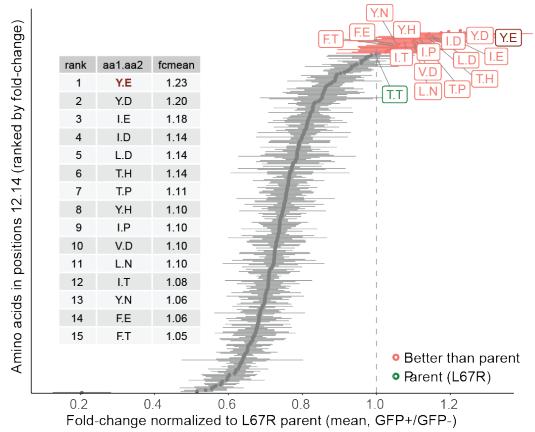


rank	aa1.aa2	fcmean
1	T.D	1.26
2	T.P	1.26
3	V.H	1.26
4	T.H	1.23
5	L.H	1.23
6	F.E	1.22
7	T.E	1.21
10	Y.D	1.20
11	Y.N	1.19
12	Y.H	1.18
13	F.T	1.14
15	Y.T	1.14
16	I.Q	1.13
17	I.E	1.13
18	I.S	1.13
22	I.Y	1.11
24	I.T	1.11

B

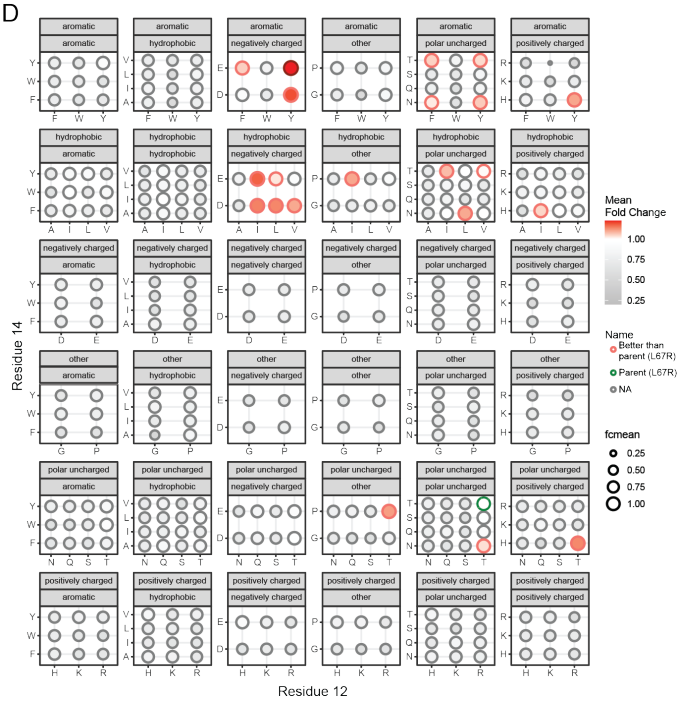


C

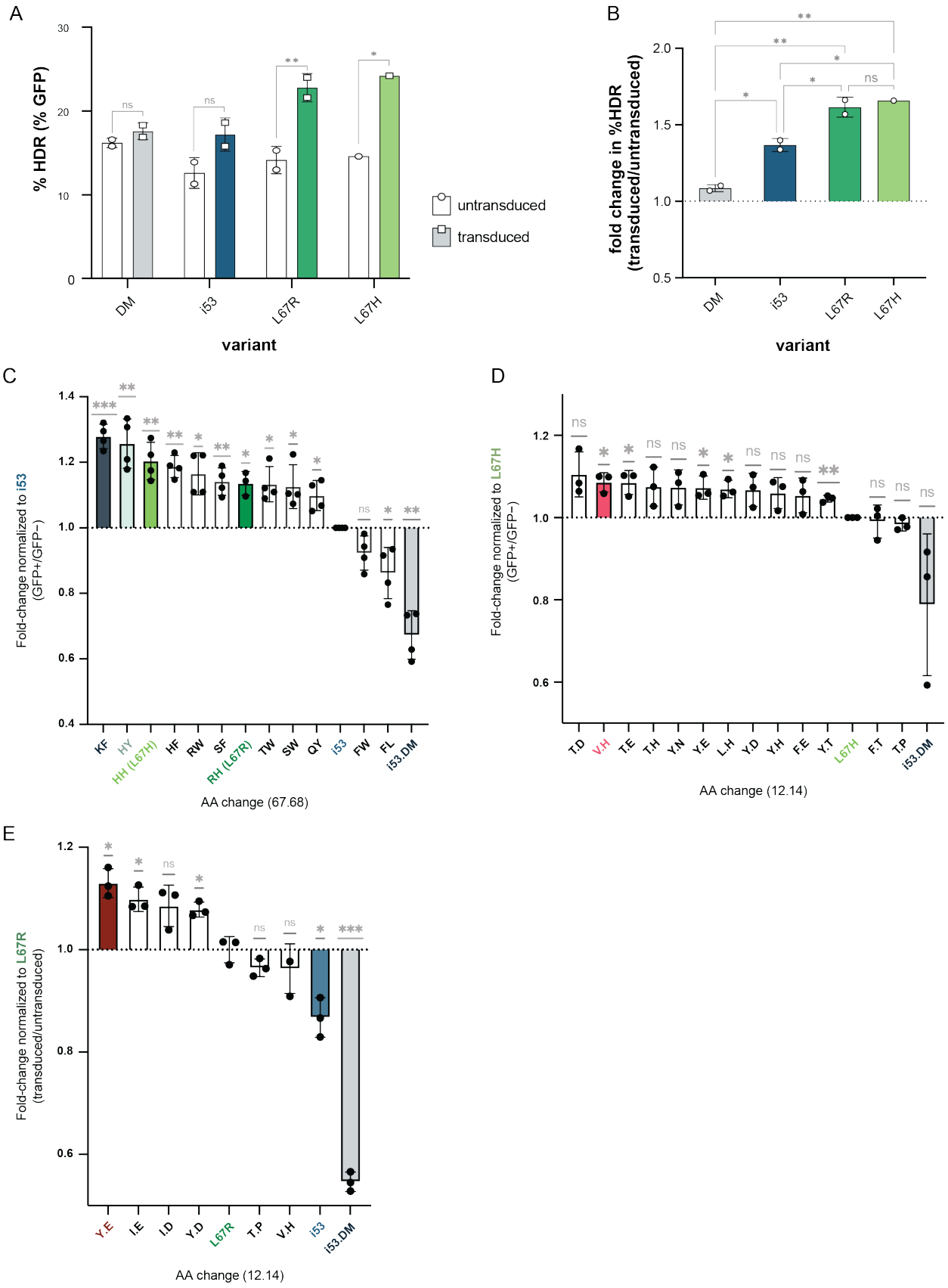


rank	aa1.aa2	fcmean
1	Y.E	1.23
2	Y.D	1.20
3	I.E	1.18
4	I.D	1.14
5	L.D	1.14
6	T.H	1.14
7	T.P	1.11
8	Y.H	1.10
9	I.P	1.10
10	V.D	1.10
11	L.N	1.10
12	I.T	1.08
13	Y.N	1.06
14	F.E	1.06
15	F.T	1.05

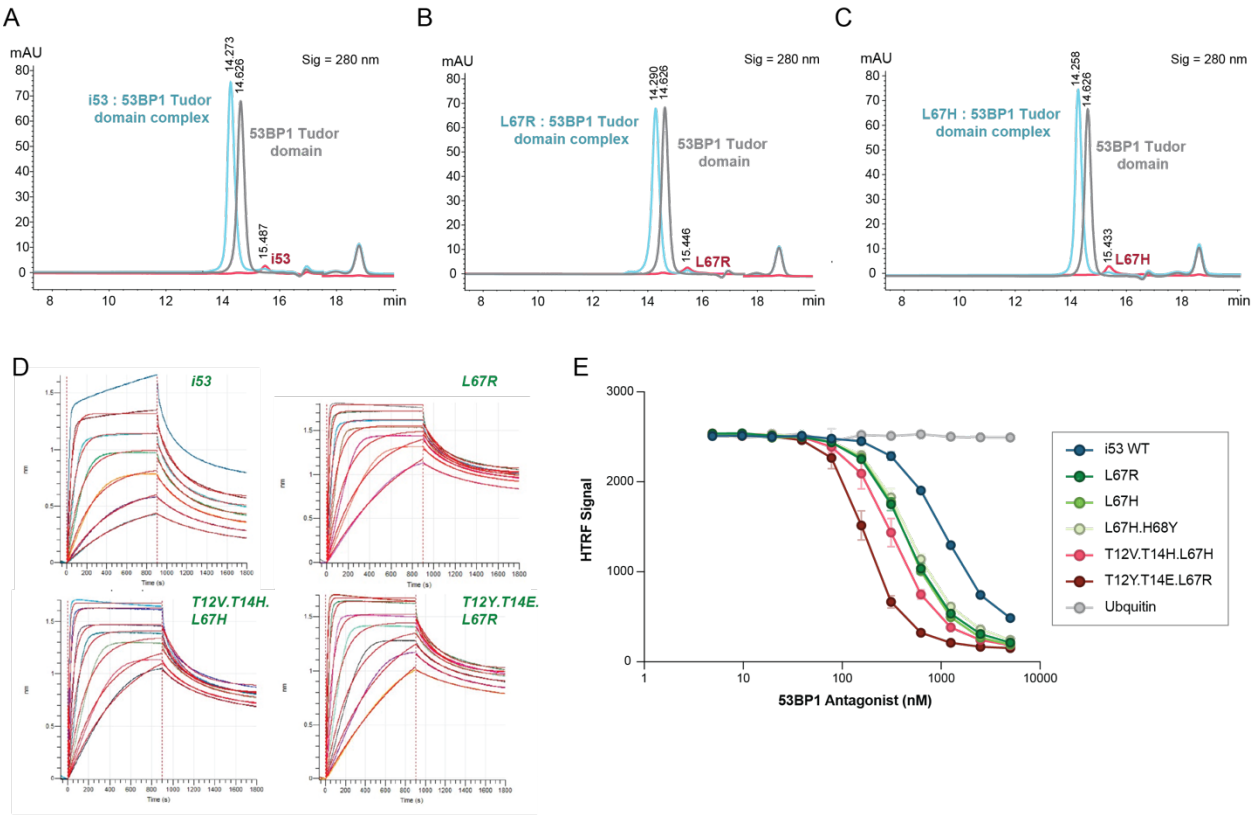
D



Supplemental Figure 1.4 **(related to Figure 1D): Screening of a combinatorial library targeting residues 12 and 14 of L67H and L67R. (A)** A combinatorial library was designed to explore all amino acid combinations (excluding cysteine and methionine, N = 324) at additional residues (T12, T14) at the 53BP1/i53 interface using L67H as the parent sequence. CD34+ HSPC cells were transduced using lentivirus packaged with the combinatorial library and edited in triplicate at the HBB locus using HBB-UbC-GFP AAV6 (MOI = 2500). Differential variant enrichment was calculated from the variant abundance in sorted mCherry+GFP+ over mCherry+GFP- populations and graphed relative to parent L67H, shown in green. Variants for which all replicates were enriched over parent are highlighted in red. (B) Dot plot representation of variant fold change enrichment as in (A), clustered by amino acid properties (amino acid variations of residues 12 and 14 shown on the x-axis and y-axis, respectively). Note amino acid properties listed for each cluster shown refer to those for residue 12 above and to residue 14 below. *N = 3* separate pooled libraries. (C) A combinatorial library was designed to explore all amino acid combinations (excluding cysteine and methionine, N = 324) at additional residues (T12, T14) at the 53BP1/i53 interface using L67R as the parent sequence. CD34+ HSPC cells were transduced using lentivirus packaged with the combinatorial library and edited in triplicate at the HBB locus using HBB-UbC-GFP AAV6 (MOI = 2500). Differential variant enrichment was calculated from the variant abundance in sorted mCherry+GFP+ over mCherry+GFP- populations and graphed relative to parent L67R, shown in green. Variants with average fold change over parent (i53) larger than 1.0 are highlighted in red. (D) Dot plot representation of variant fold change enrichment as in (C), clustered by amino acid properties (amino acid variations of residues 12 and 14 shown on the x-axis and y-axis, respectively). Note amino acid properties listed for each cluster shown refer to those for residue 12 above and to residue 14 below. *N = 3* separate pooled libraries.



***Supplemental Figure 1.5*: Validation of the top hits from the NNK (single residue) and combinatorial libraries at residues via lentiviral expression.** (A) CD34+ HSPCs were transduced to express either L67R, L67H, i53 parent or negative control i53 dead mutant (DM) and edited in duplicate at the HBB locus using HBB-UbC-GFP AAV6 (MOI = 1250). Flow cytometry was used to compare frequency of GFP+ positive cells in mCherry+ fraction expressing i53 variant to mCherry- fraction (no variant expression). *N* = two separate pooled libraries. Analysis by one-way ANOVA with post-hoc multiple comparisons analysis. *ns*= not significant; *: *p-val* < 0.05; **: *p-val* < 0.01. (B) Fold change in %GFP positive cells as in (A) showing impact of variant expression on HDR-based outcomes. *N* = two separate pooled libraries. Analysis by one-way ANOVA with post-hoc multiple comparisons analysis. *ns*= not significant; *: *p-val* < 0.05; **: *p-val* < 0.01. (C-E) Top variants identified from the various combinatorial screens were cloned individually into the lentiviral vector shown above. The resulting plasmids were either pooled together along with controls to generate small validation libraries (C, D) or used individually (E, alongside controls) to generate lentivirus. CD34+ HSPC cells were transduced with the resulting lentivirus batches and edited at the HBB locus using HBB-UbC-GFP AAV6 (MOI = 2500). Differential variant enrichment data shown in A and B was calculated using the fold change in the pooled library variant abundance in sorted mCherry+GFP+ and mCherry+GFP- populations (normalized to parents i53 and L67H, respectively). In order to calculate the normalized fold changes in %HDR in transduced over untransduced cells depicted in E, the rates of GFP integration in mCherry+ live cells (transduced, expressing the variant of interest) and mCherry- live cells (untransduced, control) were calculated for each replicate and normalized to the L67R average. *N* = 3 separate pooled libraries. Analysis by one-way ANOVA with post-hoc multiple comparisons analysis. *ns*= not significant; *: *p-val* < 0.05; **: *p-val* < 0.01; ***: *p-val* < 0.001.

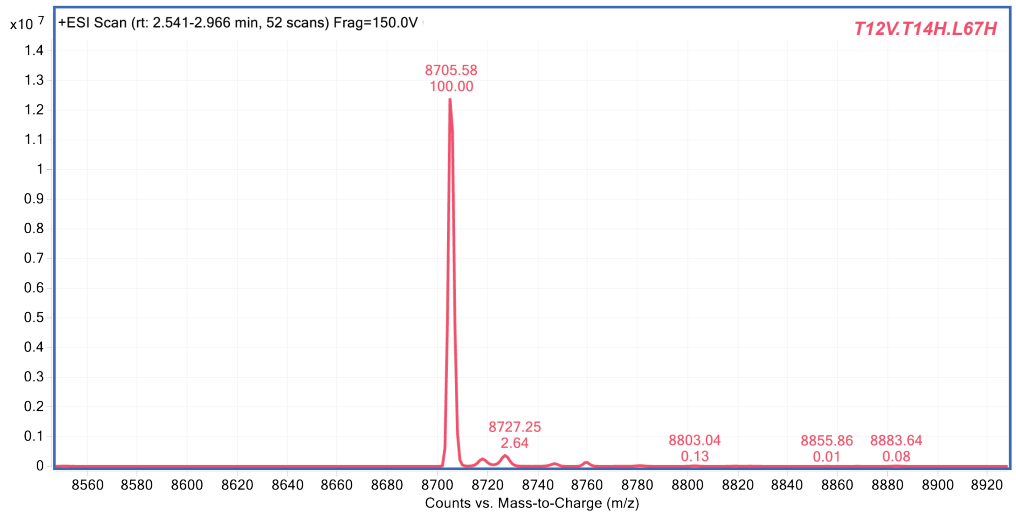


F

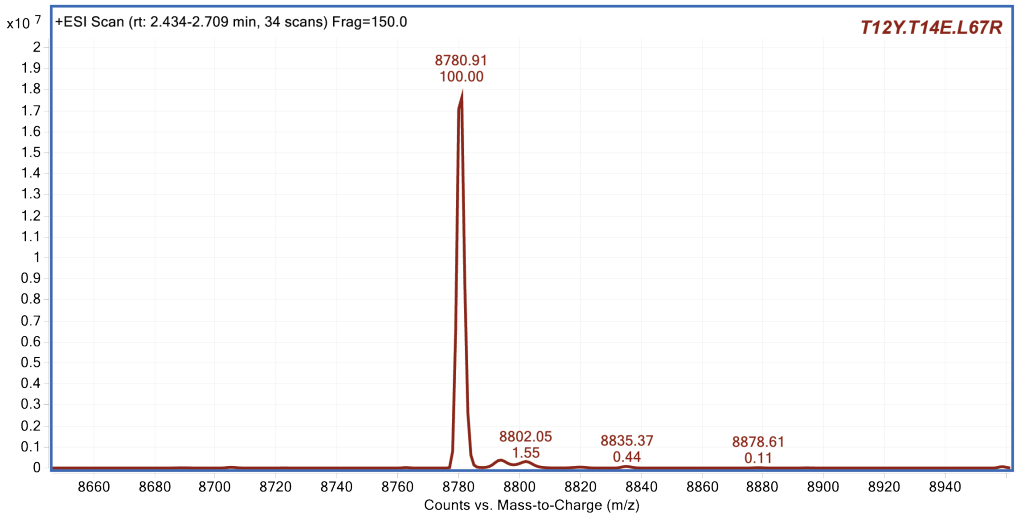
Protein name	*Fold increased affinity (BLI)*	*TR-FRET IC_{50} (nM)*
i53 (WT)	1.0	1079
L67R	2.1	458
L67H	2.2	465
L67H.H68Y	--	501
T12Y.T14E.L67R	1.2	179
T12V.T14H.L67H	2.0	338
Ubiquitin	NA	NA

***Supplemental Figure 1.6*: Validation of i53 variant binding with 53BP1 Tudor Domain.** (A-C) Example Size Exclusion Chromatography traces of i53 variants (8.6 kDa; in red) and 53BP1 Tudor domain (13.9 kDa in black) and mixed variants + Tudor domain (2-fold excess of i53; in blue) are shown: (A) WT i53, (B) L67R and (C) L67H. i53 variant retention time: ~15.5 min. 53BP1 Tudor domain retention time: 14.6 min. i53 variant and 53BP1 complex retention time: 14.3 min. (C) Immobilized 53BP1 Tudor domain binding to i53 variants by Biolayer Interferometry (BLI). (D) 53BP1 Tudor domain and i53 protein-protein interaction by TR-FRET. (E) Table listing the fold change in affinities between the K_D i53 variant proteins relative to K_D of i53 WT (as determined by BLI) and TR-FRET IC_{50}s (calculated using a non-linear 4 parameter curve fit).

A



B



***Supplemental Figure 2.1*: Mass spectrometry traces of i53 variants T12V.T14H.L67H and T12Y.T14E.L67R.** (A) T12V.T14H.L67H (expected MW = 8705.04) and (B) T12Y.T14E.L67R (expected MW = 8780.00). Y-axis quantifies spectral counts for peptide species; X-axis shows mass-to-charge ratio (m/z). Mass spectrometry confirms mutated residues at amino acid positions 12, 14 and 67.

***Supplemental Table 2.1*: Crystallography Data Table.**

Data Set	**WT**	**L67R**	**L67H**	**T12V.T14H.L67H**	**T12Y.T14E.L67R**
PDB Code	**8SVG**	**8SVH**	**8SVI**	**8SVJ**	**8T2D**
Data Collection					
Wavelength	0.95372	0.95372	0.95372	0.97939	0.97939
Space group	P21 21 21	P21 21 21	P21 21 21	P21 21 21	P21 21 21
Cell dimensions:					
a,b,c (Å)	40.37 46.86 90.32	40.46 46.96 90.04	40.38 46.87 89.73	39.98 46.92 90.57	39.510 46.85 91.12
Resolution (Å)	1.21-25.33 (1.21-1.23)	1.16-25.34 (1.16-1.18)	1.15-25.21 (1.15-1.17)	1.50-22.71 (1.50-1.53)	1.75-29.85 (1.75-1.81)
R_{symm}[a] (%)	0.05269 (0.3571)	0.08622 (0.4187)	0.08667 (0.9862)	0.05869 (0.7245)	0.06022 (1.635)
$R_{r.i.m.}$[b] (%)	0.05672 (0.3831)	0.0931 (0.4695)	0.0935 (1.101	0.06342 (0.7789)	0.0653 (1.896)
$R_{p.i.m.}$[c] (%)	0.02066 (0.1369)	0.03459 (0.2087)	0.03447 (0.4734)	0.02361 (0.2826)	0.02474 (0.9374)
Completeness (%)	96.93 (93.5)	99.8 (99.1)	98.0 (86)	99.8 (99.7)	98.89 (94.24)
Redundancy	7.5 (7.5)	7.0 (4.3)	7.0 (4.8)	7.3 (7.3)	6.6 (3.8)
I/σ	18.68 (5.5)	12.0 (3.0)	10.1 (1.6)	15.9 (2.5)	12.41 (0.59)
Wilson B factor ($Å^2$)	12.28	13.55	13.83	25.35	33.26
CC1/2	0.999 (0.952)	0.995 (0.873)	0.997 (0.558)	0.998 (0.879)	0.999 (0.386)
Refinement					
Resolution (Å)	25.3-1.21	25.34-1.16	25-1.15	22.71-1.5	29.85 -1.75
Reflections	51451	60078	59957	27947	17557
Nonhydrogen Atoms	1822	1810	1830	1695	1576
Water Molecules	241	227	187	111	38
R_{work}[d]	0.1804	0.1977	0.2055	0.2096	0.2271
R_{free}[e]	0.1892	0.2194	0.2198	0.2195	0.2673
R.m.s. deviations					
Bond lengths (Å)	0.014	0.014	0.014	0.012	0.009
Bond angles (°)	1.95	1.93	1.87	1.82	1.53
B factors ($Å^2$)					
Protein	15.16	16.31	13.49	32.43	45.14
Water	25.06	26.69	27.07	38.78	39.80
Coordinate error (Å)	0.027	0.029	0.031	0.064	0.143
Ramachandran plot[f]					
Favored (%)	98.94	98.96	98.45	97.4	97.28
Allowed (%)	1.06	1.04	1.55	2.6	2.72
Outliers (%)	0	0	0	0	0

[a] $R_{symm} = \sum|I_i - \langle I_i\rangle|/\sum I_i$, where I_i is the intensity of the ith observation and $\langle I_i\rangle$ is the mean intensity of the reflection.

[b] $R_{r.i.m.} = \sum_{hkl}[N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple observations of symmetry-related reflections.

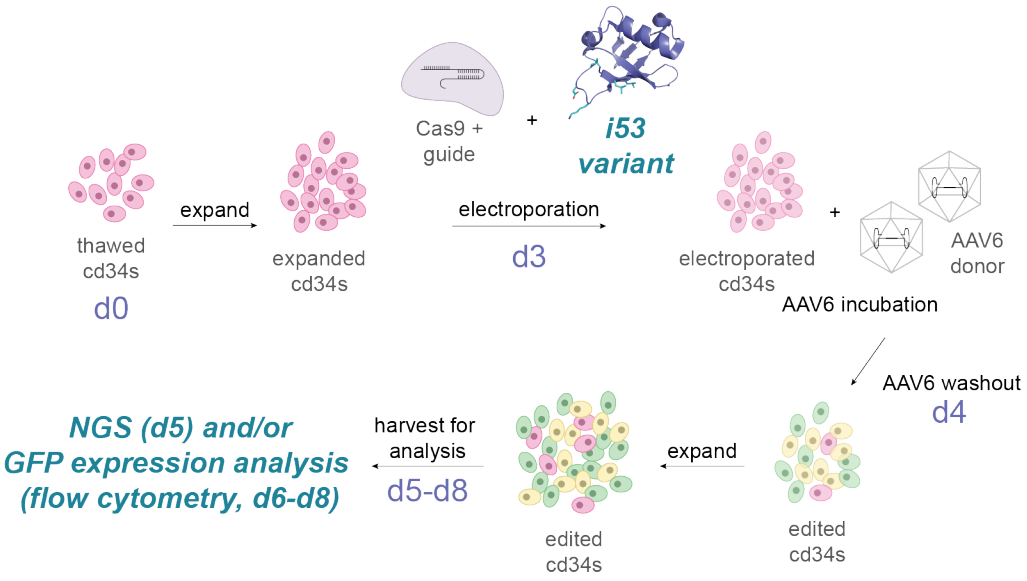
[c] $R_{p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[d] $R_{work} = \sum(||F_{obs}| - |F_{calc}|| / \sum|F_{obs}|)$

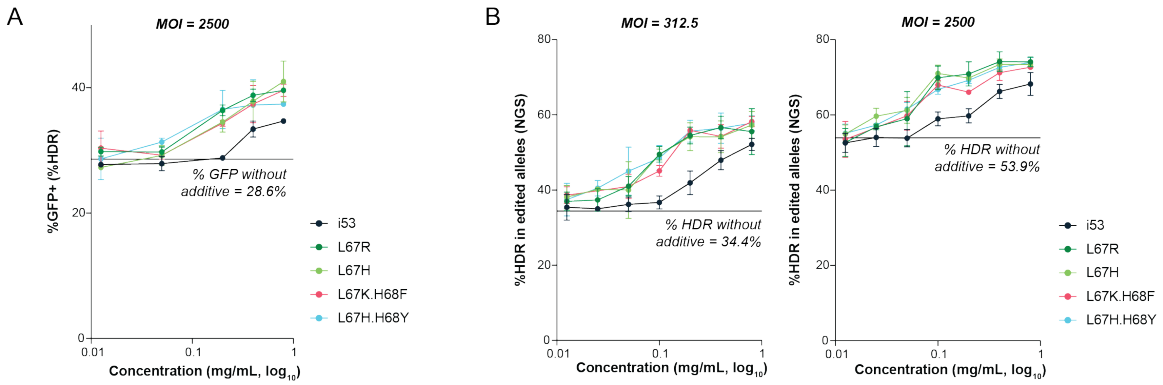
[e] R_{free} = R value for a randomly selected subset (2000 reflections) of the data that were not used for minimization of the crystallographic residual.

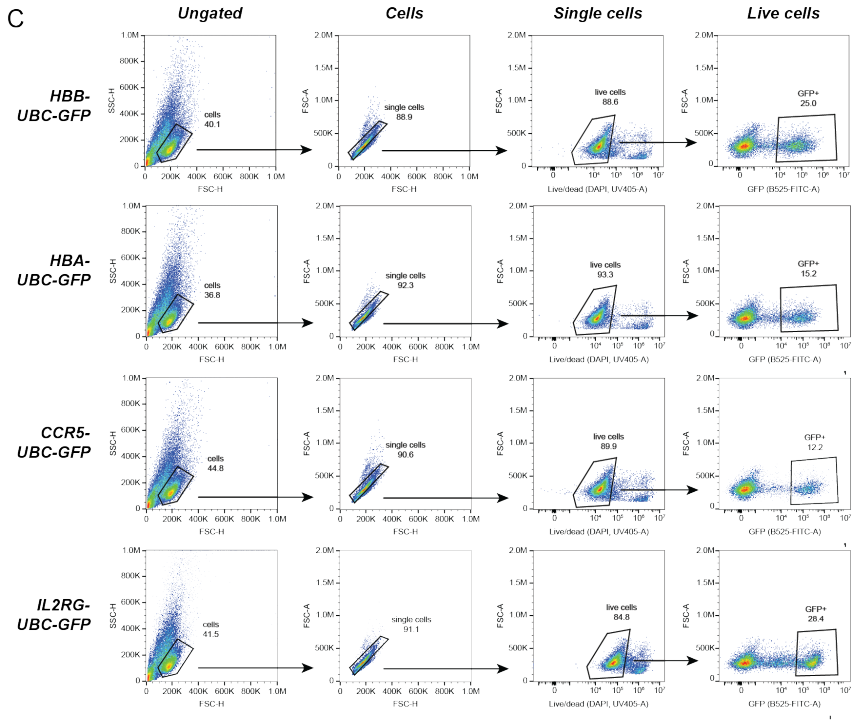
Highest resolution shell is shown in parenthesis.

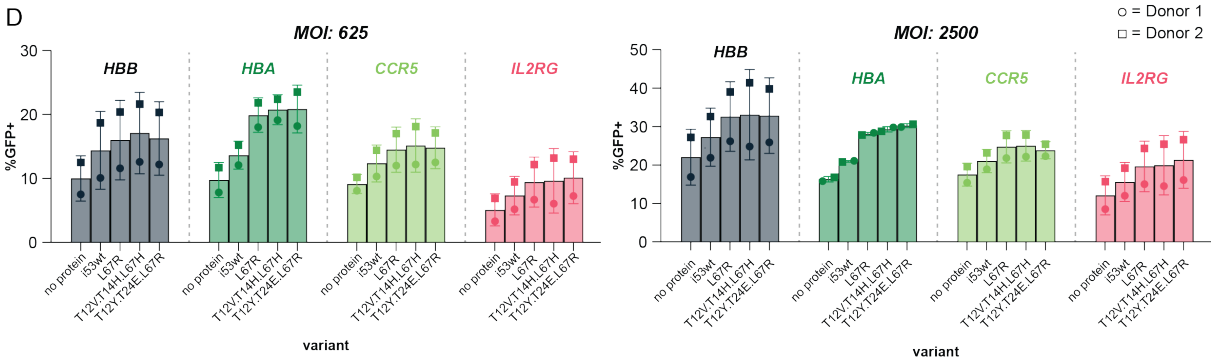
[f] Calculated with the program PROCHECK(32).



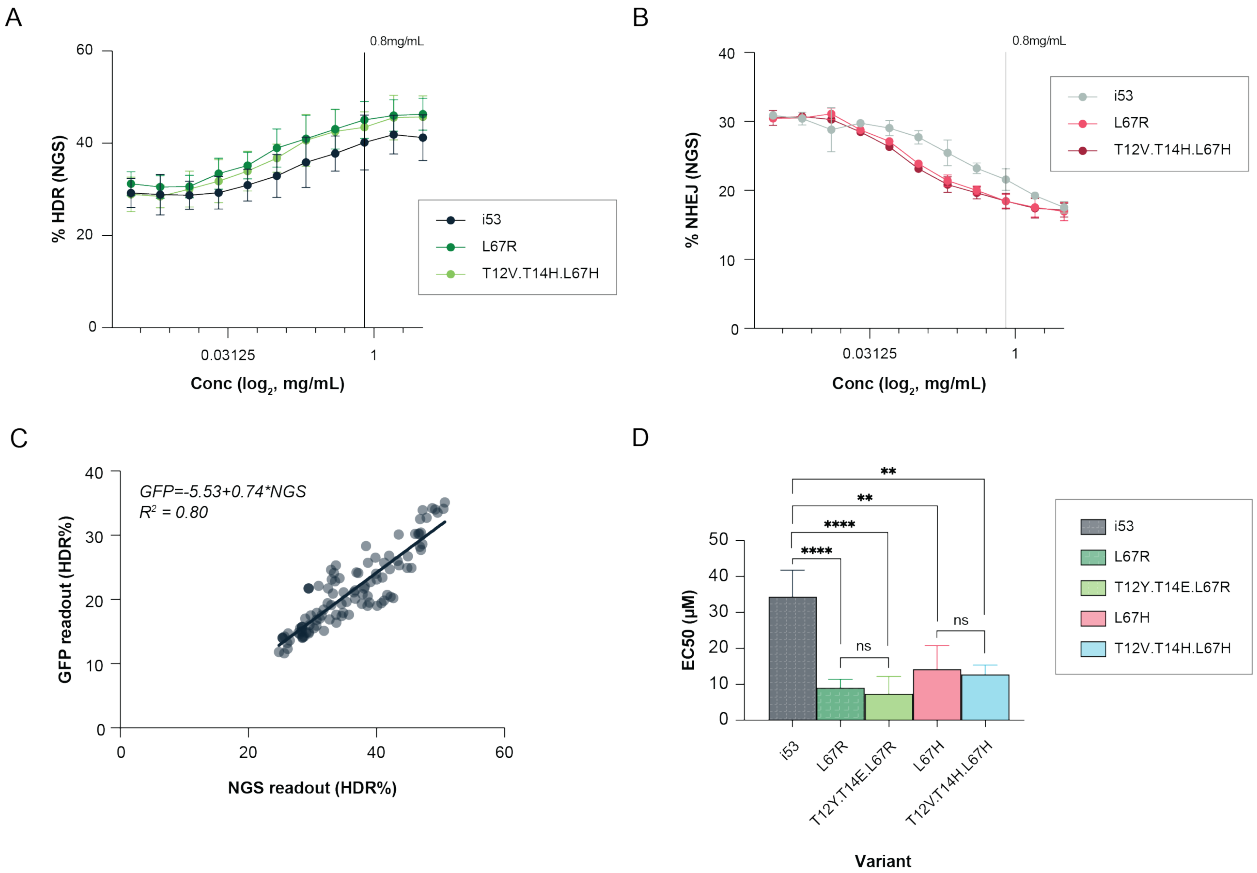
***Supplemental Figure 3.1*: Schematic detailing the procedure for CD34+ HSPC editing using Cas9 RNP and an AAV6, including incorporation of the purified i53 variant protein additives.** Editing occurs 3 days post CD34+ HSPC cell thaw; purified i53 variants are incorporated at various concentrations into the nucleofection solution containing Cas9 and guide RNA. After nucleofection, the cells are incubated in media containing various concentrations of AAV6 for 24 h and harvested for analysis 24 – 96 hours post AAV6 washout.



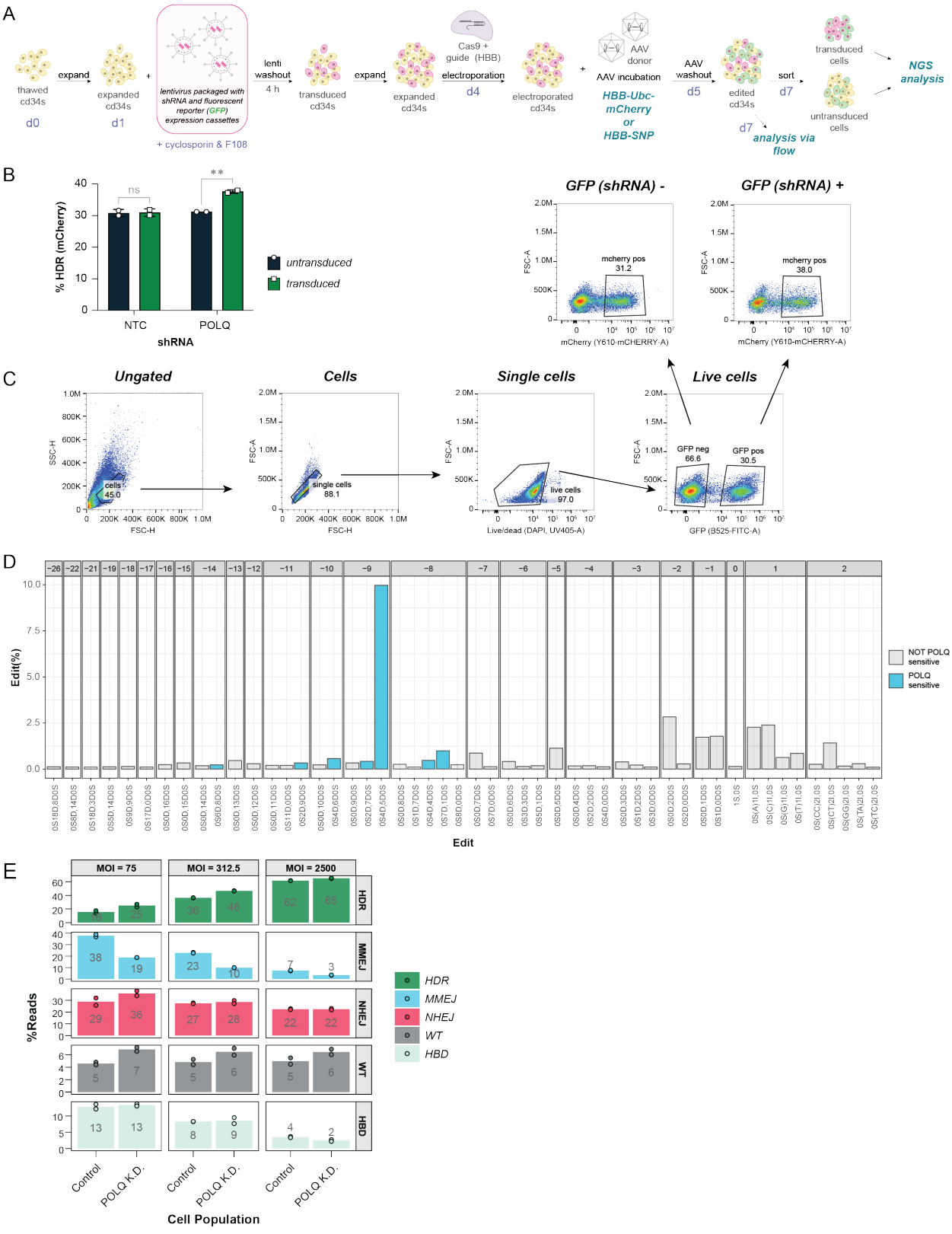




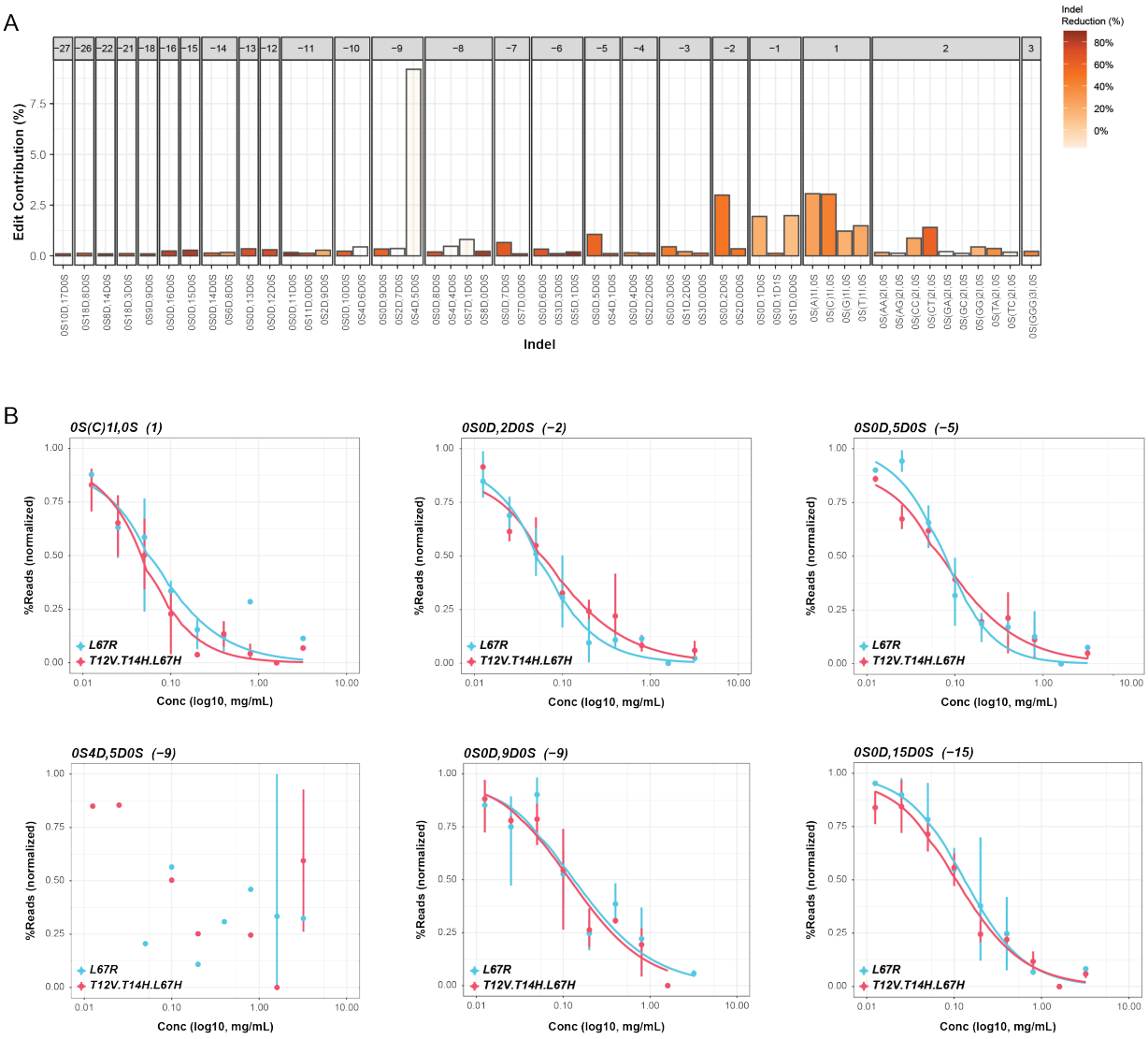
***Supplemental Figure 3.2* (relative to Figure 2A): Editing at different clinically relevant loci using GFP-encoding AAV6 and purified i53 variant proteins. (**A) %GFP positive cells in HSPCs edited with HBB-UbC-GFP AAV6 and representative purified hit i53 variants from libraries targeting residues 67 and 68. Proteins, along with an i53 control, were added to the nucleofection solutions at concentrations of 0.0125, 0.05, 0.2, 0.4, and 0.8 mg/mL (N = 2 replicates for the same CD34+ HSPC donor). (B) %HDR of edited alleles in HSPCs edited using HBB-SNP AAV6 at two different MOIs; protein variants were added to the nucleofection solutions at concentrations of 0.0125, 0.025, 0.05, 0.1, 0.2, 0.4, and 0.8 mg/mL (*N* = 4 across 3 CD34+ HSPC donors). (C) Representative flow cytometry plots showing the gating strategy to isolate Single-Live GFP+ expressing cells edited with each HBB-UbC-GFP. (C) Representative flow cytometry plots showing the gating strategy to isolate Single-Live GFP+ expressing cells edited with each AAV6. (D) GFP knock-in to CD34+ HSPC cells using GFP-encoding AAV6 targeted at multiple clinically relevant loci (*HBB*, *HBA, CCR5*, and *IL2RG*) and representative purified "hit" variants of i53 identified in the above screens (L67R, T12Y.T14E.L67R, and T12V.T14H.L67H). Protein variants, as well as a parental control WT i53, were incorporated into nucleofection solutions at concentrations of 0.4 mg/mL; post nucleofection, the cells were incubated with the AAV6 at MOIs of 625 and 2500. Cells were analyzed via flow cytometry 4 d post nucleofection; %GFP of live cells is shown. *N* = 2 separate HSPC donors.



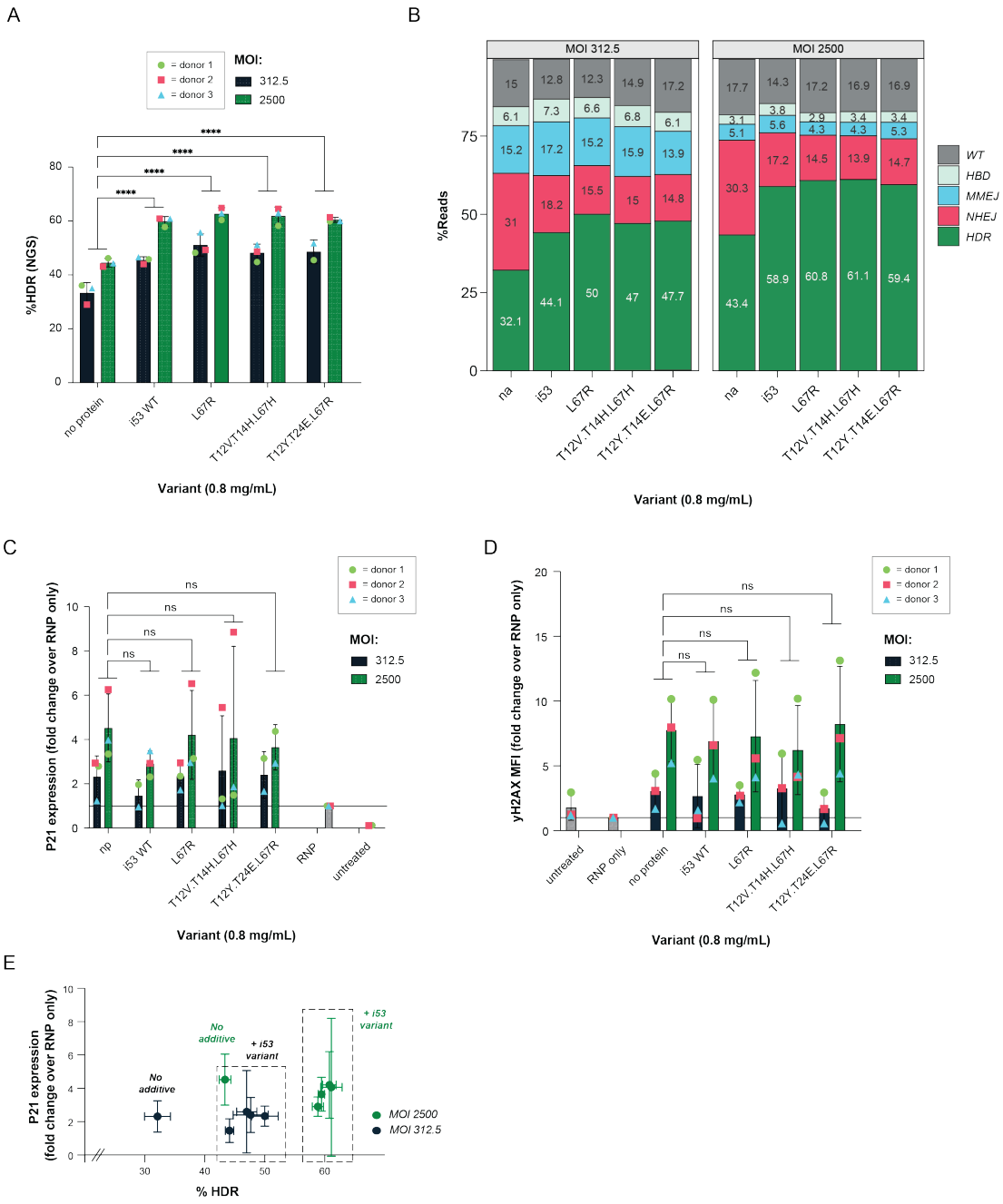
***Supplemental Figure 3.3* (related to Figure 3C): Full dose response using HBB-targeting AAV6**. Dose response curves of i53, L67R, and T12V.T14H.L67H using HBB-SNP AAV6 (MOI = 625, *N* = 3 CD34+ HSPC donors) showing as absolute HDR/NHEJ values rather than fold changes. Effects of i53 variant concentration on (A) %HDR and the corresponding effects on (B) %NHEJ of NGS reads are shown. The vertical dotted line indicates the typical working concentration (0.8 mg/mL). *N* = 3 different HSPC donors. (C) Correlation between %HDR using HBB-SNP (NGS readout) and %HDR using HBB-UbC-GFP (%GFP) across multiple editing experiments. In each experiment, cells electroporated with *HBB*-targeting RNP were split between media containing the two AAV6 conditions. Analysis by simple linear regression. (D) EC50s (shown in μM) for additional i53 variants when edited using HBB-UbC-GFP and flow cytometry readout. *N*= 3 separate HSPC donors. Four-parameter dose response curve fit, using fold change relative to no additive (0 mg/mL).



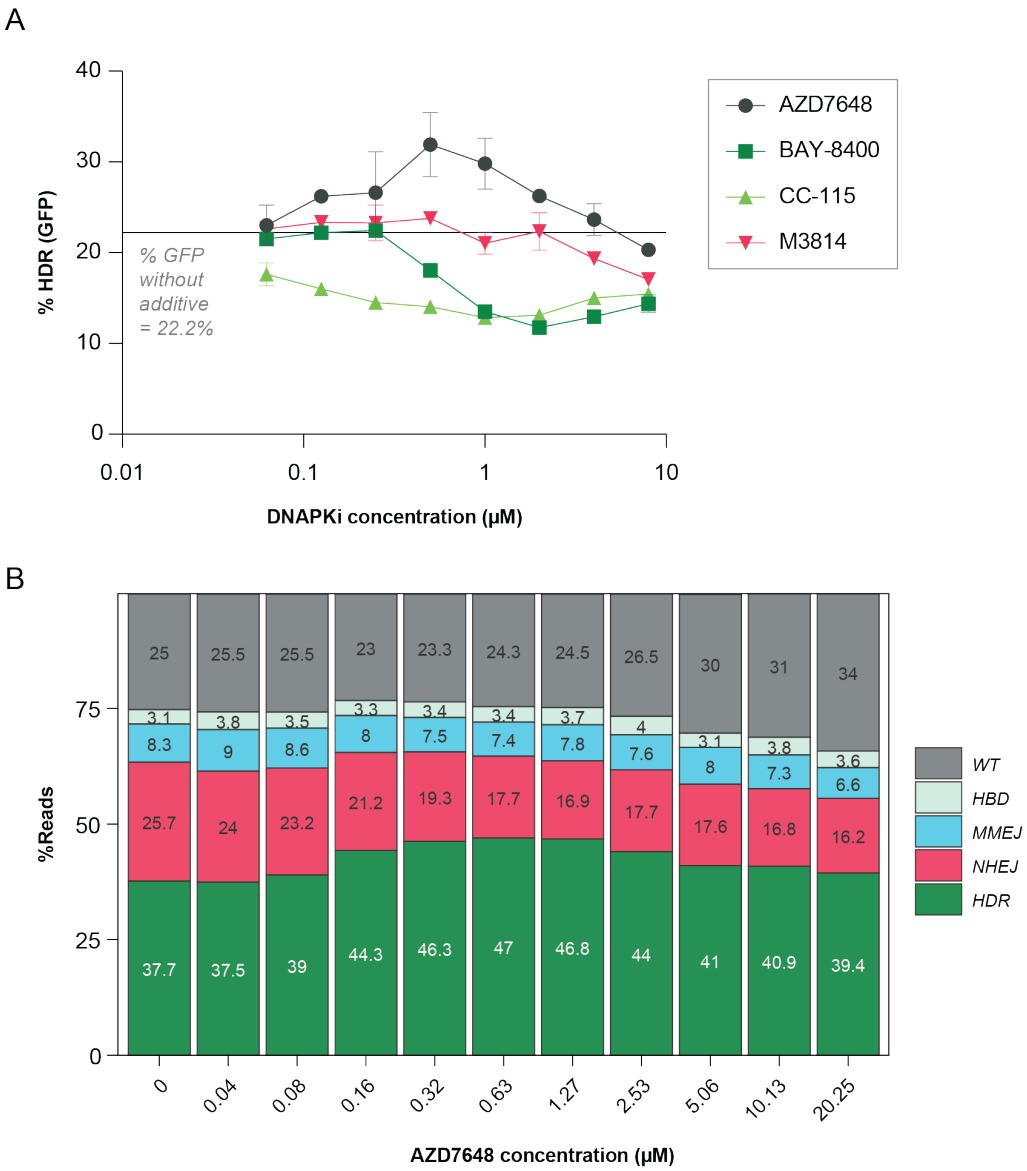
***Supplemental Figure 3.4*: Knockdown of *POLQ* elucidates MMEJ related outcomes in *HBB* locus editing.** (A) A schematic detailing the use of shRNA-encoding lentivirus (fluorescently tagged with GFP) to probe the impact of knocking down a gene of interest on HDR-based outcomes at *HBB* in CD34+ HSPCs using HBB-mCherry AAV6 or HBB-SNP DNA donors. Transduced cells edited with HBB-mCherry AAV6 were analyzed by flow cytometry; cells edited with HBB-SNP AAV6 were sorted (GFP+/GFP-) and analyzed via NGS. (B) shRNA expression was monitored by GFP expression and HDR frequency was measured by assessing mCherry positive cells in cell populations that expressed POLQ-targeting shRNA, a non-targeting control (NTC) shRNA, or no shRNA. *N* = 2 editing experiments using the same HSPC donor. Analysiss by one-way ANOVA with post-hoc pairwise comparisons and multiple testing correction. by one-way ANOVA with post-hoc pairwise comparisons and multiple testing correction. (C) Representative flow cytometry plots showing the gating strategy used to determine mCherry positive cells in transduced (GFP+) and un-transduced (GFP-) populations. (D) Individual INDEL editing outcomes at HBB and sensitivity to POLQ knockdown (HDR/WT/HBD outcomes outcomes excluded, only indels with >0.1% reads shown). Bar height represents the specific indel contribution, and fill represents if the reduction in the MMEJ knockdown was statistically significant. Using amplicon sequencing data. *N = 2* separate editing experiments using the same HSPC donor. Deemed significant if *p-val* < 0.05 using one-way ANOVA with post-hoc pairwise testing and multiple testing correction. (E) Editing outcome distribution when implementing MMEJ classification. HBB editing outcomes sensitive to POLQ knockdown are designated “MMEJ”, whereas indels that did not change in MMEJ knockdown are classified as “NHEJ”. The impact of POLQ knockdown on MMEJ edits, as well as the other categories of edits, are shown.



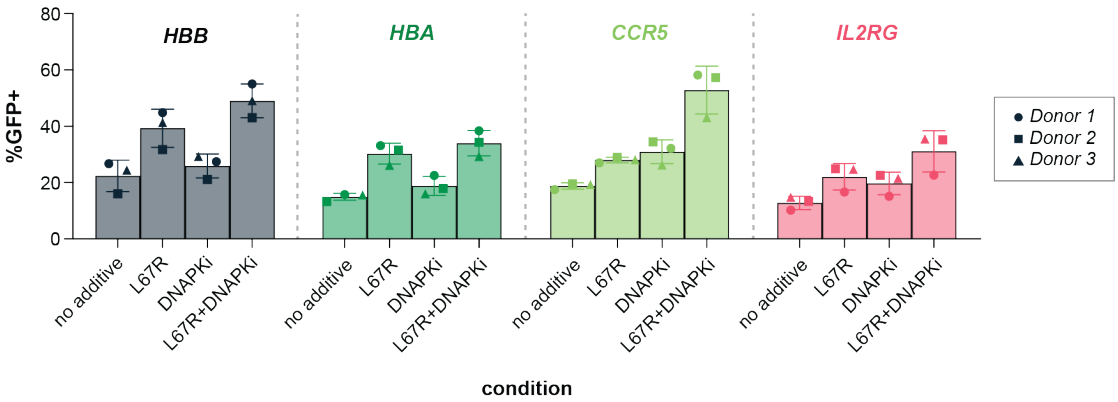
***Supplemental Figure 3.5*: Identification of 53BP1i-sensitive INDELs at the *HBB* locus cutsite. (A) Plot of the contribution of individual INDEL outcomes.** Bar shading represents normalized reduction of the prevalence of each indel when i53 variants are added for editing (HDR/WT/HBD excluded, only indels with >0.1% reads shown). White fill represent edits that did not result in a dose-dependent reduction of contribution when i53 variants were used. *N* = 3 different HSPC donors; each one of them with two different i53 variants. An individual dose-response curve (four parameters, nonlinear) was fit for each variant and only those with significant association (*p-val* < 0.01) were categorized as i53 responsive. (B) Sample plots of the contribution of specific indels at increasing concentrations of two of the i53 variants identified in this study, along with fitted dose-response curves. An MMEJ edit (-9, bottom left) is shown as an example of a non-i53 responsive edit). *N* = 3 separate donors edited.



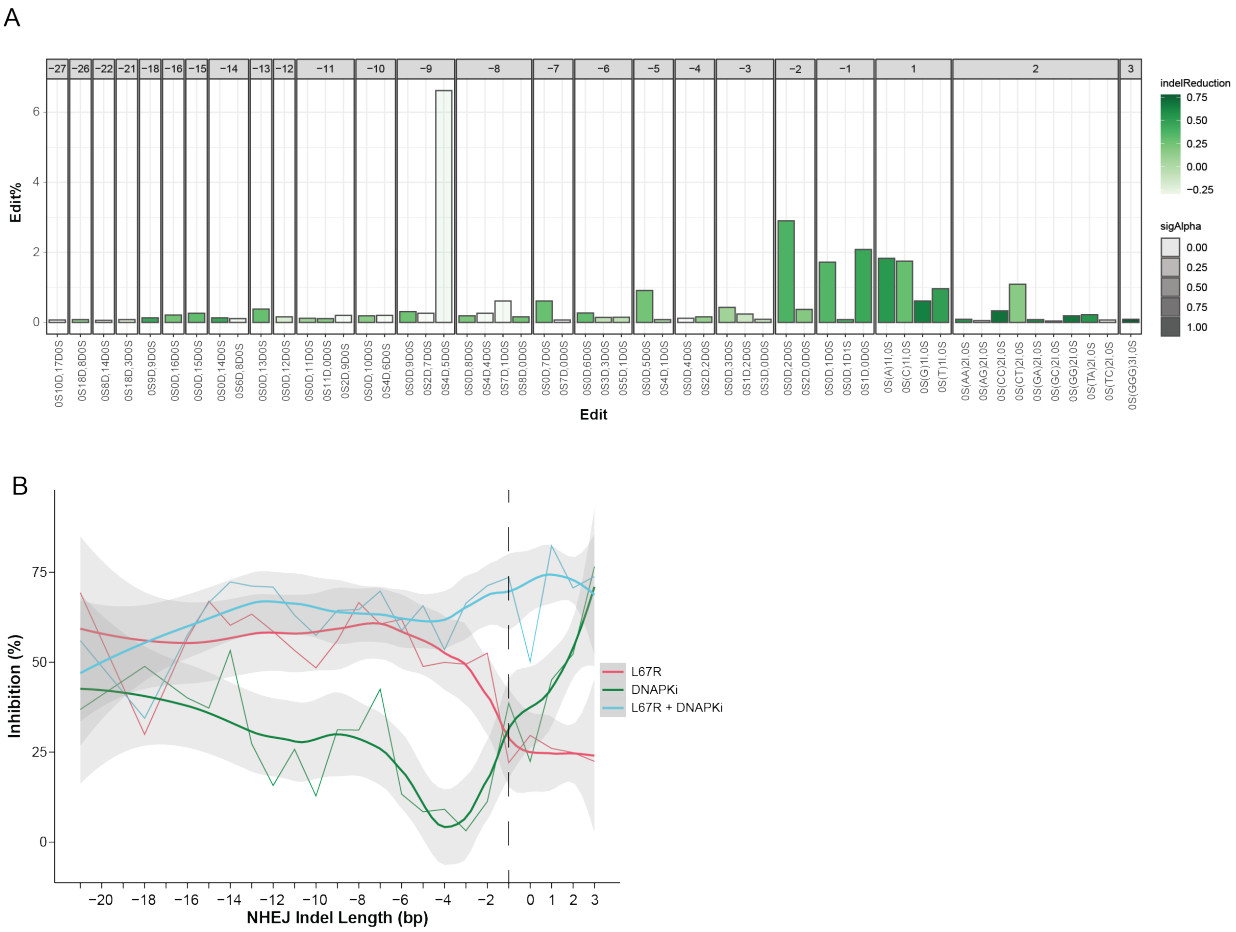
***Supplemental Figure 3.6* (related to Figure 3D and 3E): Editing of HSPCs using HBB-SNP AAV6 and purified i53 variant proteins.** (A-B) HSPCs edited using HBB-SNP AAV6 and purified variants of i53 (MOI = 312.5 and 2500, *N* = 3). (A) %HDR and (B) averaged %reads for the different *HBB* editing outcomes as determined by NGS analysis. *N = 3* different HSPC donors. Analysis by two-way ANOVA with post-hoc pairwise comparison and multiple testing correction. Only showing the variant effect. ***: *p-val* < .0001. (C-D) Induction of the DNA Damage Response (DDR) as measured by (C) expression of P21 and (D) phosphorylation of histone H2AX (yH2AX) 24 hours post nucleofection. *N=3* different HSPC donors. Analysis by two-way ANOVA with post-hoc pairwise comparison and multiple testing correction. Only showing the variant effect. *ns*: not significant. (E) %HDR relative to P21 expression with and without the addition of the i53 variants.



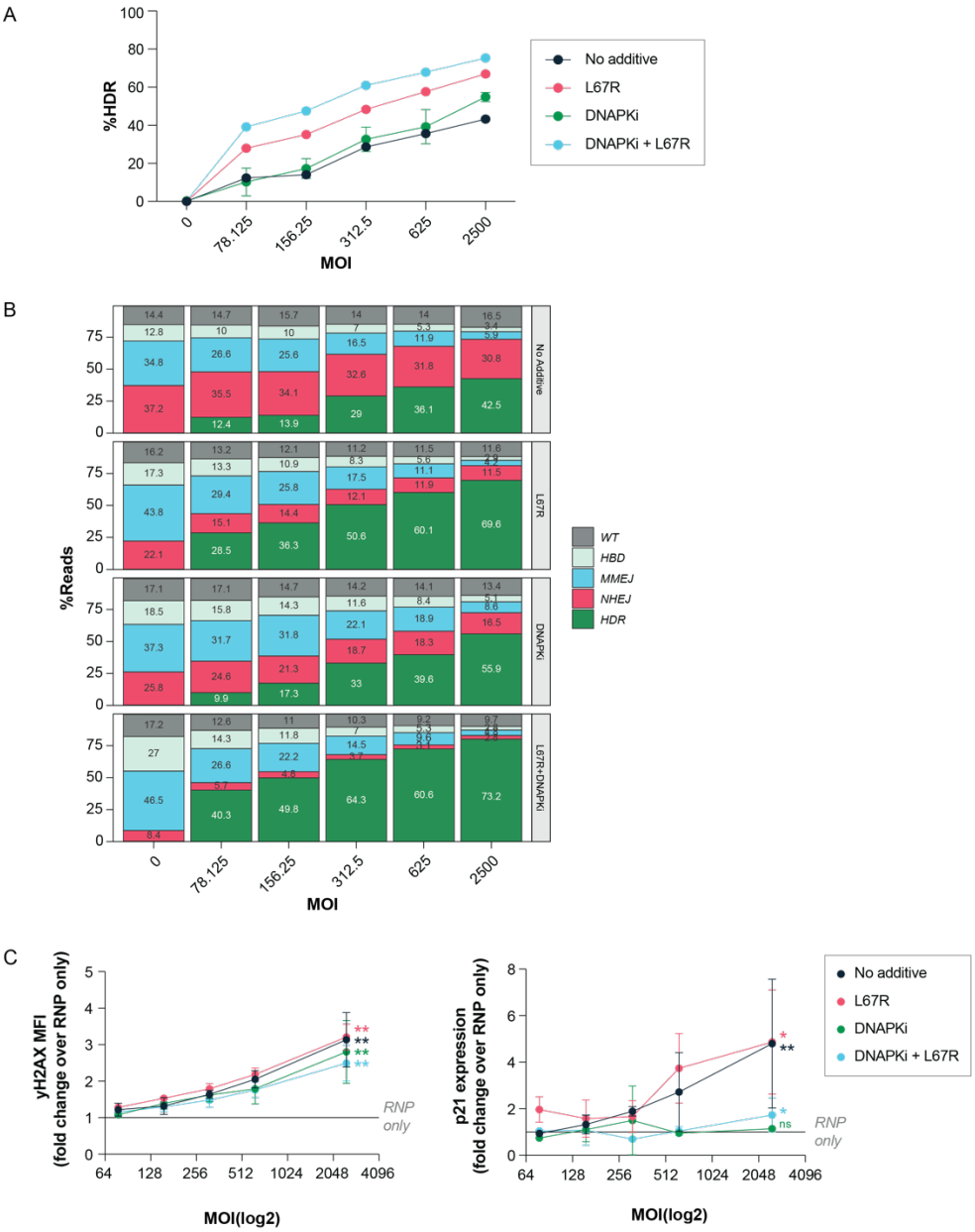
***Supplemental Figure 4.1*: DNAPKi dose response curves.** (A) Dose response curves when adding different DNAPKcs-targeting small molecules to an HSPC editing protocols using HBB-UbC-GFP AAV6 (MOI = 2500, *N* = 2). (B) Dose response curve of the effect of different types editing outcomes at HBB as determined by NGS when AZD7648 is added to HBB-SNP AAV6-containing media post nucleofection. (MOI = 2500, *N* = 3).



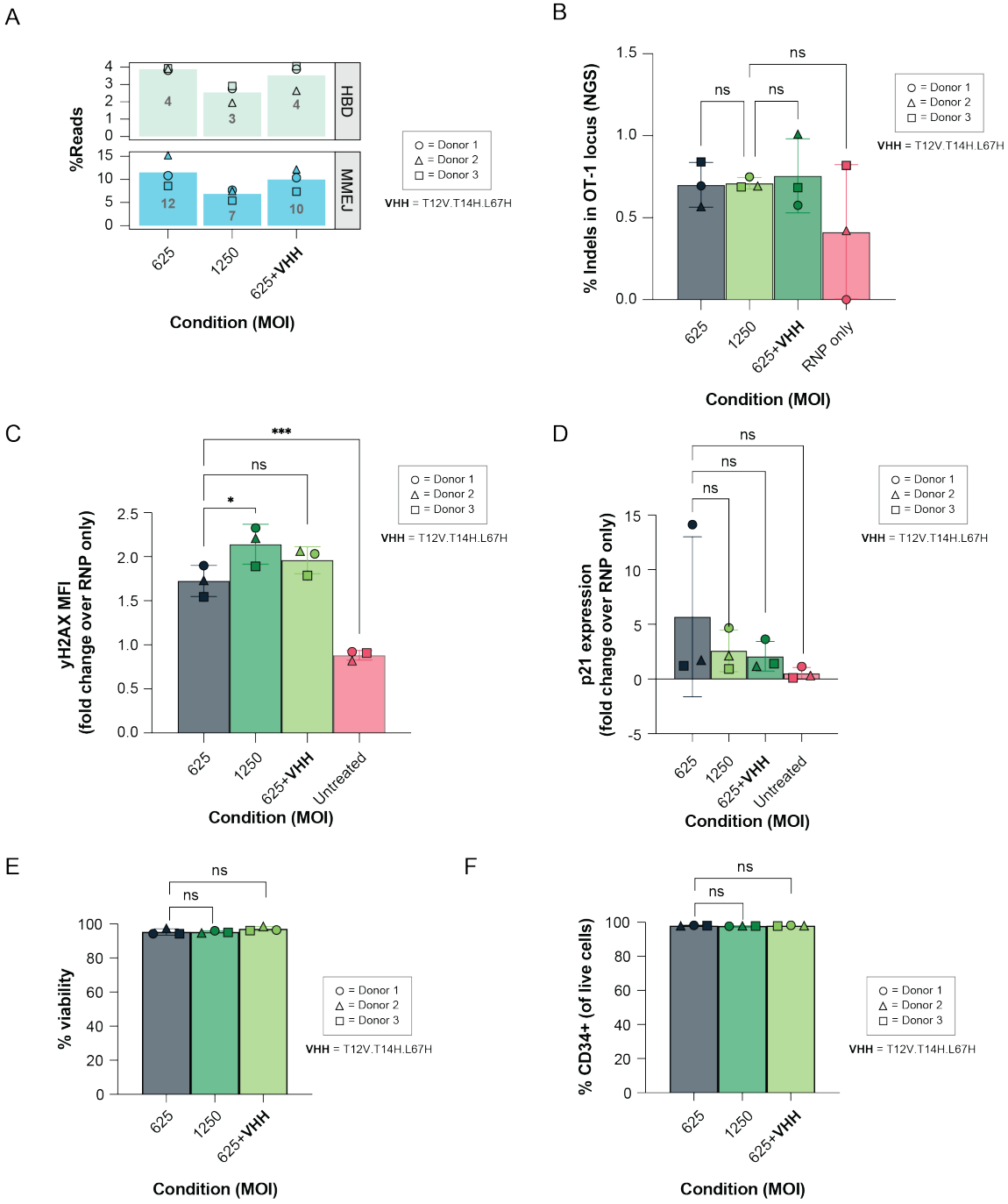
***Supplemental Figure 4.2* (related to Figure 4A): Editing at different clinically relevant loci using GFP-encoding AAV6, purified i53 variant protein, and AZD7648.** %GFP-expressing cells (%HDR) when L67R (0.8 mg/mL) is incorporated to an HSPC editing protocol for GFP knock-ins; post editing, cells were resuspended in media containing AAV6 targeted at *HBB*, *HBA*, *CCR5*, and *IL2RG* (MOI = 2500) with and without the addition of a DNAPKi (AZD7648, 0.5 μM). Cells were analyzed via flow cytometry 4 d post nucleofection; %GFP of live cells is shown. *N* = 3 independent HSPC donors.



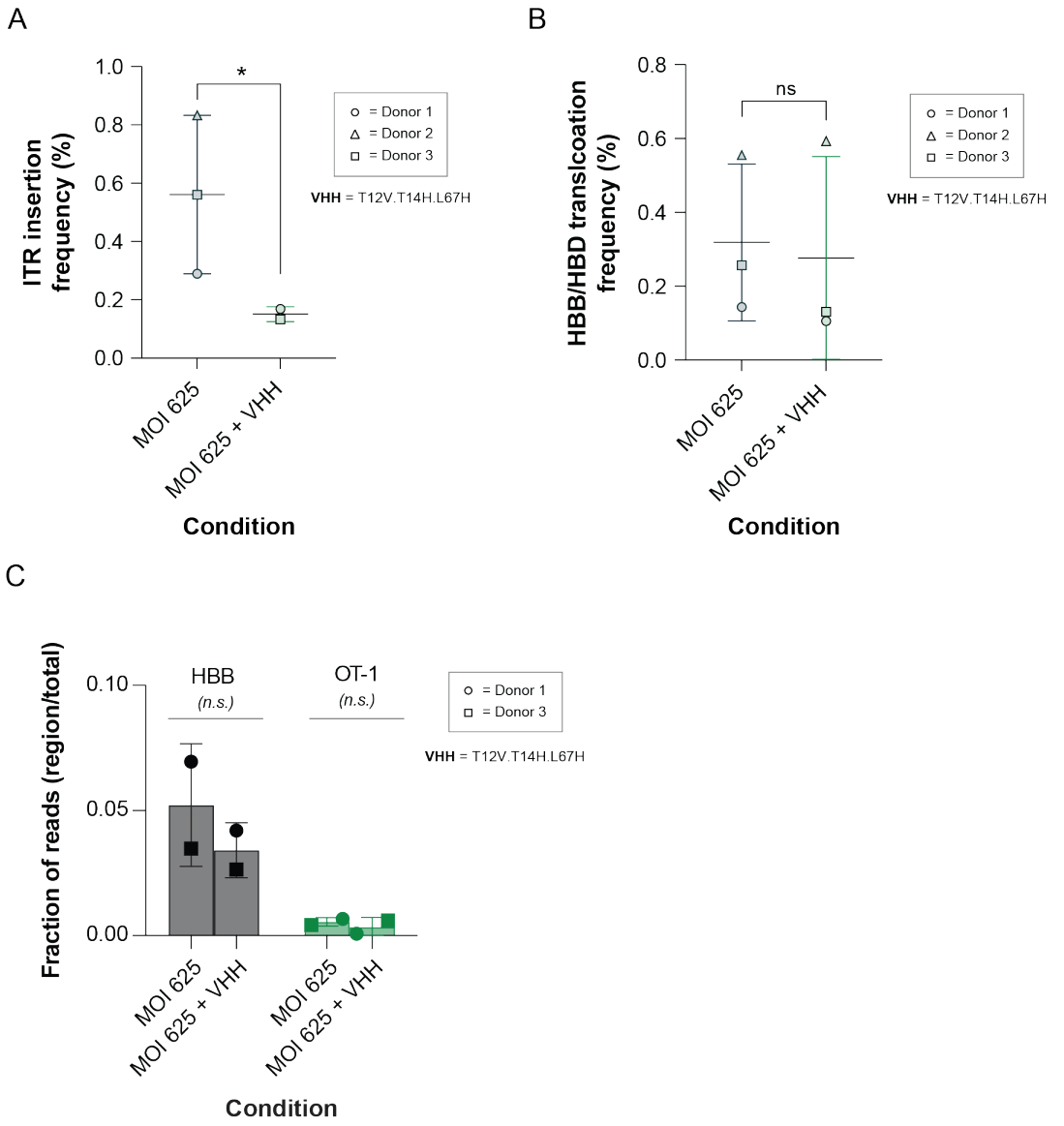
Supplemental Figure 4.3 **(related to Figure 4B): DNAPKi-sensitive INDELs and a comparison to 53BP1i-sensitive INDELs.** (A) Contribution of individual INDEL editing outcomes at *HBB* and sensitivity to DNAPKi AZD7648 (HDR/WT/HBD excluded, only indels with >0.1% reads shown). Bar shading represents reduction when DNAPKi small molecule is used during editing. White fill represents edits that did not result in a dose-dependent reduction of contribution when i53 variants were used. *N* = 3 different HSPC donors. An individual dose-response curve (four parameters, nonlinear) was fit for each variant and only those with significant association (*p-val* < 0.01) were categorized as i53 variant responsive. (B) Differential effects of NHEJ inhibitors (L67R and AZD7648) alone and in combination on the %inhibition individual NHEJ-derived INDELs at *HBB* (MMEJ excluded), at different indel lengths. A vertical dashed line indicates the indel length at which DNAPKi small molecule yields better inhibition than 53BP1 inhibitor L67R. *N* = 3 different HSPC donors.



***Supplemental Figure 4.4* (related to Figure 4C): MOI titration of the HBB-SNP AAV6 donor with L67R and AZD7648.** (A) %HDR in cells treated with no additive, L67R (0.8 mg/mL), AZD7648 (0.5 μM), or both L67R and AZD7648 (N = 2 CD34+ HSPC donors). (B) Impact of MOI on the different types of editing outcomes as determined by NGS for both no additive and L67R only conditions. Stacked bars represent the contribution for different repair outcome categories. *N* = 3 separate HSPC donors. (C) Induction of the DNA Damage Response (DDR) as measured by expression of P21 and phosphorylation of histone H2AX (γH2AX) 24 hours post nucleofection and the differential effects for each editing condition across MOIs. *N* = 3 different HSPC donors. Significance is provided by slope of linear regression being different from zero. *ns: not significant*, *: *p-val* < 0.05; **: *p-val* < 0.01.



***Supplemental Figure 5.1* (Related to Figure 5A): Medium scale production run editing outcomes and cell health metrics.** (A) HBD and MMEJ editing outcomes in CD34+ HSPCs from 3 donors that were edited with Cas9 RNP and HBB-SNP AAV6 at medium scale (~200 M HSPCs per condition for each donor). *N*=3 different HSPC donors. (B) %INDELs at known off-target site OT-1. *N=3* different HSPC donors. Analysis by one-way ANOVA with post-hoc pairwise comparisons and multiple testing adjustment. N*s*= not significant. (C-D) Induction of DNA Damage Response (DDR) was measured by (C) yH2AX phosphorylation and (D) P21 expression 24 hours post nucleofection. (E) Viability and (F) %CD34+ of cells post cryopreservation. For all panels, *N=3* different HSPC donors. Analysis by one-way ANOVA with post-hoc pairwise comparisons and multiple testing adjustment. *ns*= not significant; *: p-value<0.05; ***: p-value < 0.001.



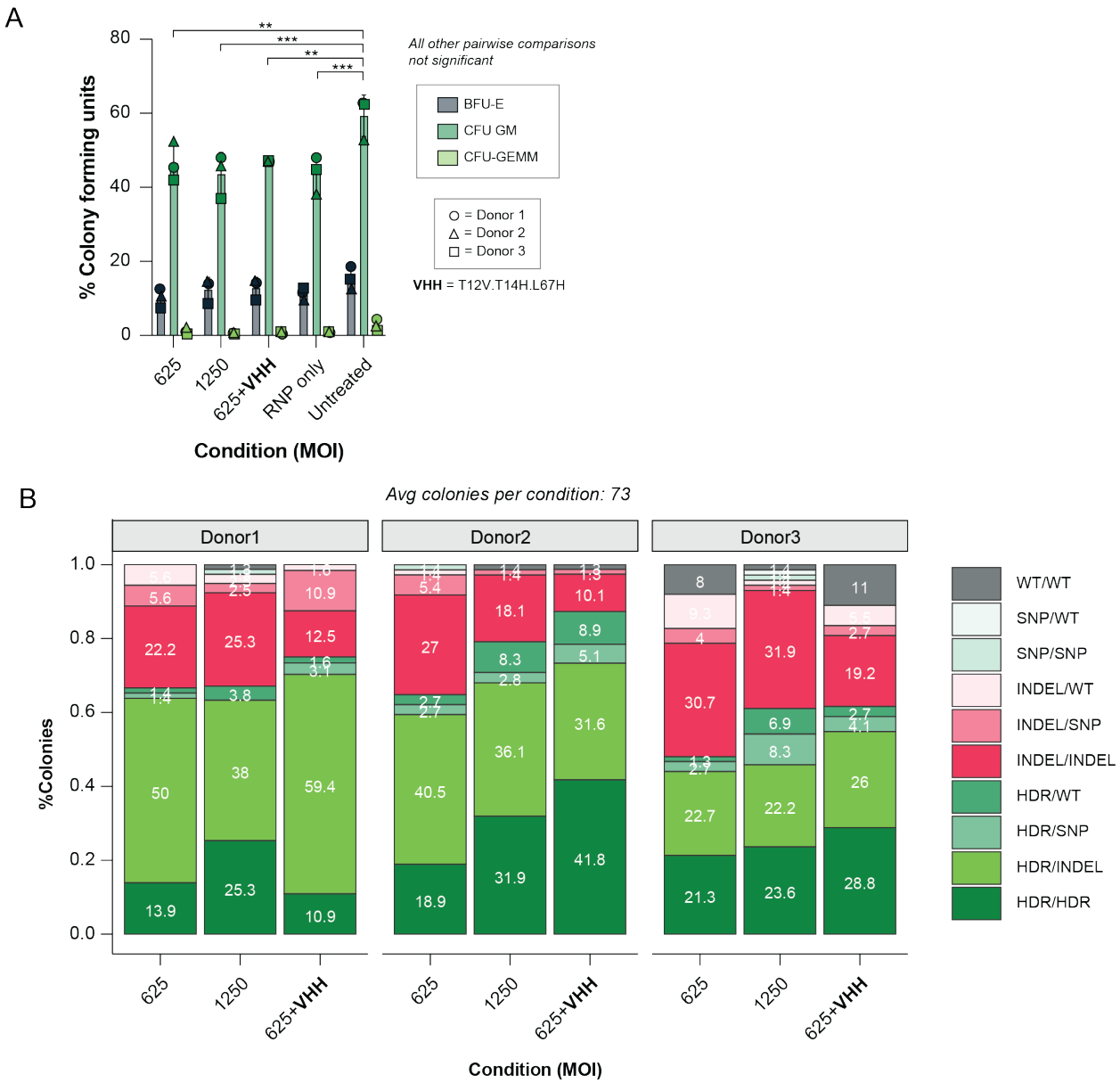
***Supplemental Figure 5.2*: Translocation-sequencing and Guide-Seq results**. (A) Summed frequency of AAV ITR sequence insertions identified at HBB and OT1 by translocation sequencing in bulk edited HSPC cells from medium scale production runs (MOI 625 and MOI 625 + T12V.T14H.L67H conditions only). Note that ITR integration was not identified in donor 2 using the MOI 625 + VHH editing condition. P-value = 0.0482 by two-sided unpaired t-test, setting missing conditions as 0. (B) Summed frequency of translocations identified between the HBB and HBD loci by translocation sequencing in bulk edited HSPC cells from medium scale production runs (MOI 625 and MOI 625 + T12V.T14H.L67H conditions only). Not statistically significant by two-sided unpaired t-test, setting missing conditions as 0. (B) Guide-seq characterization of cut sites for *HBB* locus targeting gene editing, with and without i53 variants used as additives. All the identified cutsites around *HBB* cutsite (on-target) and OT-1 (off-target) were aggregated into single columns. These were the only significant insertion sites not present in the untreated samples. *N*=2 separate donors. All comparisons are non-significant by paired t-test with multiple comparisons adjustment.

***Supplemental Table 5.1*: Karyotyping of HSPC under different conditions**. In all conditions the evaluation was of normal karyotype, with some non-clonal, low level chromosomal gain/loss which was deemed normal for cell culture. *N* = 3 separate HSPC donors for all conditions; 100 spreads/condition.

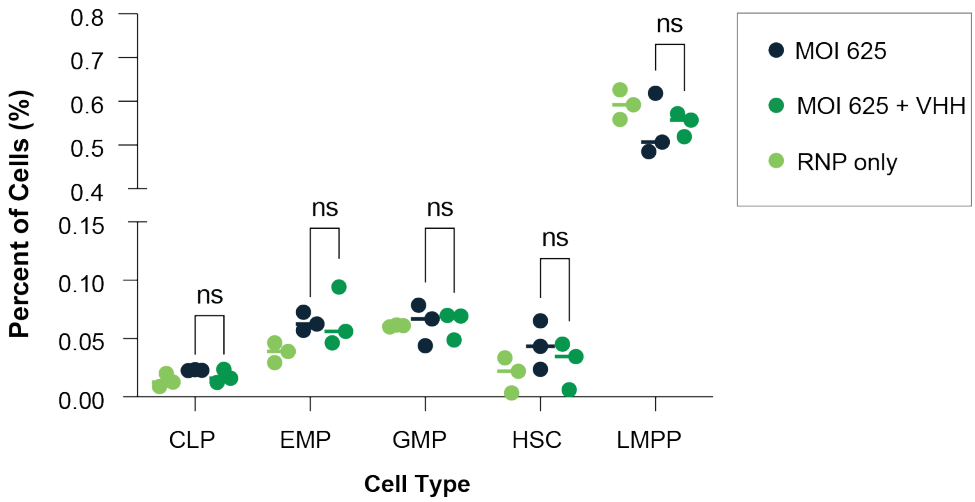
Donor	Treatment	Total cells	Case notes	Number of Normal karyotype cells (46,XY/46,XX)	Chromosome gain/loss	Chr11 gain/loss (HBB)	Chr9 gain/loss (OT-1)
1	Untreated	100	Normal Karyotype	90	10	0	0
1	MOI 625	100	Normal Karyotype	82	18	0	1
1	MOI 625 + VHH	100	Normal Karyotype	80	20	1	1
2	Untreated	100	Normal Karyotype	89	11	0	1
2	MOI 625	100	Normal Karyotype	93	7	3	0
2	MOI 625 + VHH	100	Normal Karyotype	96	4	0	1
3	Untreated	100	Normal Karyotype	91	9	0	1
3	MOI 625	100	Normal Karyotype	88	12	1	0
3	MOI 625 + VHH	100	Normal Karyotype	84	16	1	1

***Supplemental Table 5.2*: Guide-seq characterization of cut sites for *HBB* locus targeting gene editing, with and without i53 variants used as additives.** This list includes all the matched integration records, whose left and right break points (BPs) have >= 1 UMI reads.

Sample ID	L_BP_chr	L_Peak_Pos	R_Peak_Pos	Distance	Int_Dir	Region_reads	Region/Total(E-6)	Comment	Gene
donor 1	chr11	5226983	5226984	0	+	25018	65824.37	On-Target	HBB
donor 1	chr9	101833600	101833601	0	+	2523	6638.22	OT-1	NA
donor 1	chr11	5226979	5226983	3	-	1397	3675.62	On-Target	HBB
donor 1 + VHH	chr11	5226983	5226984	0	+	23827	35629.74	On-Target	HBB
donor 1 + VHH	chr9	101833600	101833604	3	+	535	800.01	OT-1	NA
donor 1 + VHH	chr11	5226985	5226985	-1	-	4230	6325.34	On-Target	HBB
donor 3	chr11	5226983	5226984	0	+	16724	27364.18	On-Target	HBB
donor 3	chr9	101833600	101833603	2	+	2619	4285.27	OT-1	NA
donor 3	chr11	5226979	5226985	5	-	4562	7464.44	On-Target	HBB
donor 3 + VHH	chr9	101833600	101833601	0	-	3736	6112.93	OT-1	NA
donor 3 + VHH	chr11	5226983	5226984	0	+	25119	24988.98	On-Target	HBB
donor 3 + VHH	chr11	5226984	5226985	0	-	1415	1407.68	On-Target	HBB



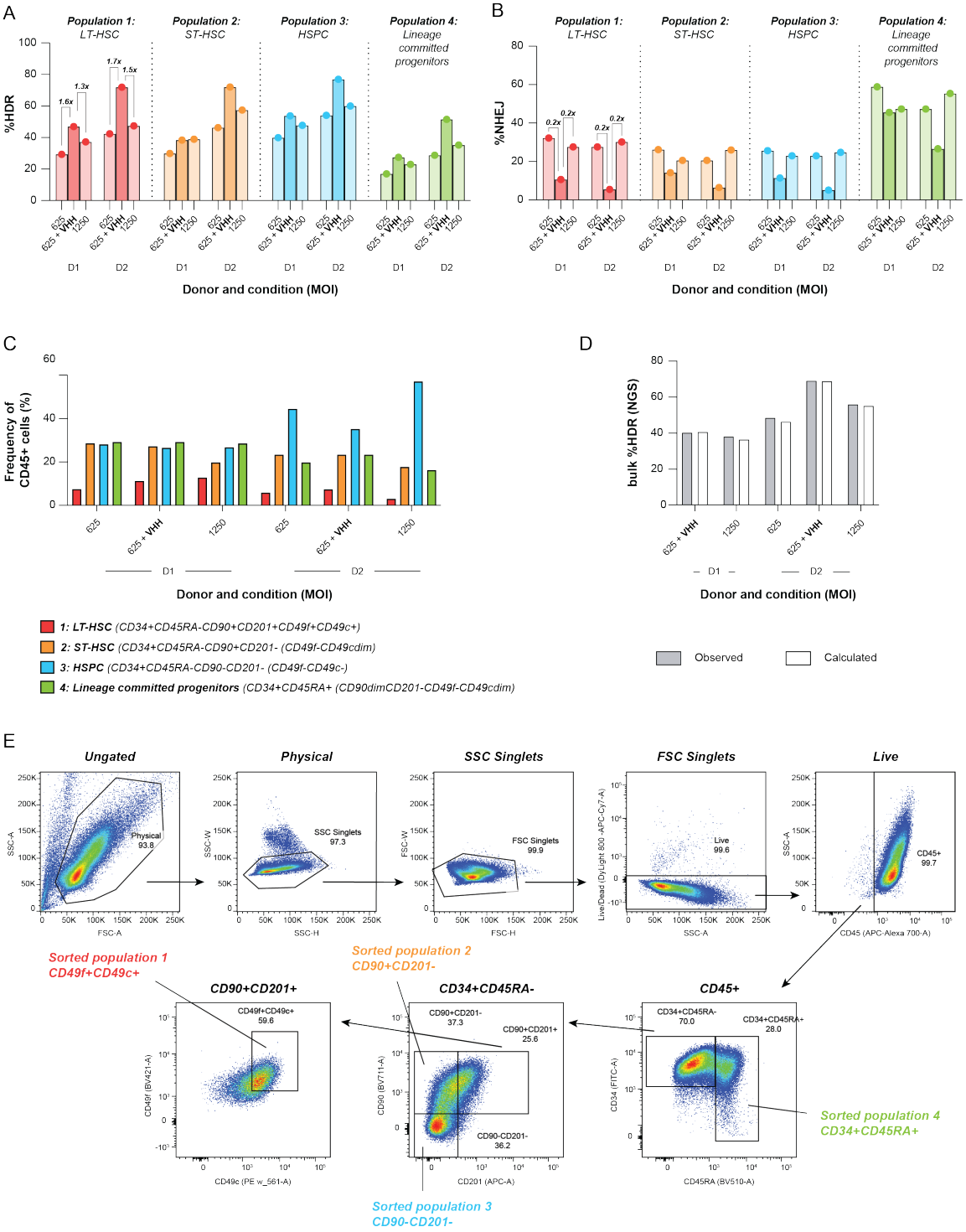
***Supplemental Figure 5.3* (related to Figure 5B): CFU counts and CFU sequencing results for medium scale production run editing.** (A) HSPC health and repopulation capacity was assessed by colony forming unit (CFU) recovery. CFUs were counted and stratified by type using a STEMvision colony counter instrument (StemCell Technologies): BFU-E (burst forming unit-erythroid), GM (granulocyte-macrophage progenitor), GEMM (multipotent progenitor granulocyte, erythrocyte, monocyte, megakaryocyte). *N*=3 different HSPC donors. Analysis by two-way ANOVA with post-hoc pairwise correction (Tukey correction). Only showing condition effect. (B) Breakdown of genotypes from the sequencing of individual colonies from each condition and donor. VHH = T12V.T14H.L67H (0.8 mg/mL).



***Supplemental Figure 5.4*: i53 variant addition does not alter the cell mixture composition.** Single cell RNA sequencing (scRNA-seq) characterization of cell type composition from cell pools, 5 days post gene editing. For simplicity, only cell types that are relevant to engraftment (HSC and lineage-committed progenitors) are shown. Data was analyzed using separate t-test (two-tailed, unpaired) and adjusting for multiple comparisons. ns: p-value >0.1. Cell type acronyms: CLP: common lymphoid progenitor; EMP: erythroid-megakaryocyte progenitor; GMP: granulocyte-monocyte progenitor; HSC: hematopoietic stem cell; LMPP: lymphoid-primed multipotential progenitor. Cell counts for all cell types can be found in Supplemental Table 5.2 VHH = T12V.T14H.L67H (0.8 mg/mL).

***Supplemental Table 5.2*: scRNA-Seq cell numbers.** VHH = T12V.T14H.L67H (0.8 mg/mL).

Cell type	MOI 625			MOI 625 + VHH			RNP only (no AAV)			Untreated		
	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
ASDC	0	0	1	0	2	4	0	3	1	0	0	0
BaEoMa	39	58	150	65	83	99	21	139	105	16	20	54
CD14 Mono	1	0	9	2	0	0	0	0	1	5	6	35
CD16 Mono	0	0	0	0	0	0	0	0	0	6	1	6
CD4 Effector	0	0	0	0	0	0	0	0	0	0	1	0
CD4 Memory	1	0	2	1	0	0	0	5	1	3	7	167
CD4 Naive	1	0	7	1	10	19	25	5	0	222	264	355
CD8 Effector_1	0	0	0	0	0	0	0	0	0	1	0	1
CD8 Memory	0	0	0	0	0	0	0	0	0	0	1	0
cDC2	0	0	0	0	0	3	0	0	1	0	0	0
CLP	65	68	139	50	62	96	22	88	92	307	174	377
Early Eryth	376	590	1154	537	894	862	365	1419	906	164	169	289
EMP	163	219	369	381	179	228	72	267	213	759	1062	1182
GMP	125	237	395	198	268	284	151	419	276	33	36	87
HSC	186	71	256	183	24	140	54	23	154	2452	2303	2078
Late Eryth	0	0	1	0	0	0	1	0	0	10	4	2
LMPP	1764	1460	2998	2311	2156	2106	1530	4062	2566	818	1024	429
Macrophage	0	0	0	0	0	0	0	0	0	0	0	5
Memory B	5	2	5	2	1	7	0	1	6	30	2	29
Naive B	0	0	0	0	0	0	1	0	0	2	2	8
NK	0	0	0	0	0	0	0	0	0	2	3	0
pDC	1	5	7	1	1	0	0	11	1	1	2	0
Plasma	42	59	97	79	60	25	21	54	40	120	63	106
Platelet	0	0	0	0	0	0	0	0	0	4	0	51
pre B	4	1	4	7	2	3	6	0	0	74	15	32
pre-mDC	5	28	65	16	26	49	1	50	45	0	0	1
pre-pDC	0	7	32	4	14	43	2	27	22	0	7	4
pro B	3	3	11	17	5	15	14	16	27	93	30	23
Prog Mk	71	202	207	184	82	73	152	262	137	12	16	22
transitional B	0	0	1	3	0	0	3	3	0	11	3	1
Total	2852	3010	5910	4042	3869	4056	2441	6854	4594	5145	5215	5344



***Supplemental Figure 5.5* (related to Figure 5C): LT-HSC-enriched HSC subpopulation sort and editing outcomes for the medium-scale production runs.** (A) %HDR and (B) %NHEJ in HSPC subpopulations sorted from bulk edited cells as in Fig S4.2 (donor 1 and 2 only). VHH = T12V.T14H.L67H (0.8 mg/mL). D1 = donor 1. D2 = donor 2. (C) Observed frequency of sorted subpopulations for each donor and condition. (D) Calculated (from %HDR and frequency of each subpopulation) and observed %HDR in bulk edited cells. (E) Representative flow cytometry plots showing the sort gating strategy to isolate Single-Live populations of HSCs using designated markers.