

# CRISPR-Cas9 engineering of the *RAG2* locus via complete coding sequence replacement for therapeutic applications

Received: 8 February 2023

Accepted: 28 September 2023

Published online: 27 October 2023

 Check for updates

Daniel Allen<sup>1,5</sup>, Orli Knop<sup>1,5</sup>, Bryan Itkowitz<sup>1,5</sup>, Nechama Kalter<sup>1</sup>, Michael Rosenberg<sup>1</sup>, Ortal Iancu<sup>1</sup>, Katia Beider<sup>2</sup>, Yu Nee Lee<sup>3,4</sup>, Arnon Nagler<sup>2,3</sup>, Raz Somech<sup>3,4</sup> & Ayal Hendel<sup>1</sup>✉

*RAG2*-SCID is a primary immunodeficiency caused by mutations in *Recombination-activating gene 2 (RAG2)*, a gene intimately involved in the process of lymphocyte maturation and function. ex-vivo manipulation of a patient's own hematopoietic stem and progenitor cells (HSPCs) using CRISPR-Cas9/rAAV6 gene editing could provide a therapeutic alternative to the only current treatment, allogeneic hematopoietic stem cell transplantation (HSCT). Here we show an innovative *RAG2* correction strategy that replaces the entire endogenous coding sequence (CDS) for the purpose of preserving the critical endogenous spatiotemporal gene regulation and locus architecture. Expression of the corrective transgene leads to successful development into CD3<sup>+</sup>TCRαβ<sup>+</sup> and CD3<sup>+</sup>TCRγδ<sup>+</sup> T cells and promotes the establishment of highly diverse TRB and TRG repertoires in an in-vitro T-cell differentiation platform. Thus, our proof-of-concept study holds promise for safer gene therapy techniques of tightly regulated genes.

Severe combined immunodeficiency (SCID) is a group of multiple rare monogenic disorders characterized by defects in both cellular and humoral adaptive immunity. Patients are born healthy and due to being extremely susceptible to pathogens, they present with recurrent infections early in life which if left untreated can be fatal<sup>1–3</sup>. The *Recombination-activating genes 1 and 2 (RAG1 and RAG2)*, respectively) are tightly linked and have convergent transcriptional orientations on chromosome 11 separated by ~12 kb. The *RAG* genes encode proteins that, when complexed together, commence the lymphoid-specific variable (V), diversity (D), and joining (J) gene [V(D)J] recombination process by catalyzing DNA double-strand breaks (DSBs) at the recombination signal sequences (RSSs) which flank the V, D, and J gene segments<sup>4</sup>. V(D)J recombination is a critical step in the maturation of T and B cells as it is responsible for the generation of a diverse repertoire of T- and B-cell receptors (TCR and BCR, respectively). Thus, patients

with disease-causing variants in the *RAG* genes typically present with the complete absence or significant reduction of T and B cells and the T-B-NK<sup>+</sup> immune phenotype<sup>2,5–9</sup>. V(D)J recombination has three main mechanisms of regulation: 1) lineage specificity, namely BCR and TCR gene rearrangement occurs in B cells and T cells, respectively; 2) immunoglobulin heavy-chain rearrangement occurs before immunoglobulin light chain; and 3) allelic exclusion, namely once a T or B cell rearranges its receptor locus, only one functional allele will be expressed in that cell. In addition to these general mechanisms, the transcription of the *RAG1* and *RAG2* genes is regulated by numerous highly-conserved, lineage-specific, cis-acting sequences surrounding their respective CDSs which control the spatial genomic organization inside the locus<sup>10–15</sup>. The *RAG* genes are expressed exclusively in the G<sub>0</sub>/G<sub>1</sub> stage and display a tightly linked genomic organization with regulated expression during specific phases of T and B lymphocyte

<sup>1</sup>Institute of Nanotechnology and Advanced Materials, The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 5290002, Israel. <sup>2</sup>The Division of Hematology and Bone Marrow Transplantation, Chaim Sheba Medical Center, Tel-Hashomer, Ramat Gan 5266202, Israel. <sup>3</sup>Sackler Faculty of Medicine, Tel Aviv University, 6997801 Tel Aviv, Israel. <sup>4</sup>Pediatric Department A and the Immunology Service, Jeffrey Modell Foundation Center, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Ramat Gan 5266202, Israel. <sup>5</sup>These authors contributed equally: Daniel Allen, Orli Knop, Bryan Itkowitz. ✉e-mail: [ayal.hendel@biu.ac.il](mailto:ayal.hendel@biu.ac.il)

development. More specifically, two limited waves of *RAG1* and *RAG2* expression are necessary for Ig heavy and light chain rearrangements after which their expression is promptly terminated<sup>16</sup>. This process is orchestrated by a plethora of transcription factors and machinery that complex together with cis-regulatory elements and promoter regions in the *RAG1/2* locus. Together, they create a chromatin hub that acts as a super-enhancer for the expression of the *RAG* genes<sup>12,14,17,18</sup>. This formation, as well as the chromatin structure and 3D architecture, are crucial to ensuring that the *RAG1/2* genes are only expressed during the requisite developmental window<sup>19</sup>. Overexpression or expression of the *RAG* genes outside of this precise window can result in genomic instability and lymphocyte malignancy through the formation of translocations and/or deletions in cancer-causing genes<sup>11,12,19–23</sup>. Unsuccessful termination of *RAG1* and *RAG2* expression is also associated with atypical thymus development, an aberrant lymphatic system, and immunodeficiency<sup>24</sup>.

Currently, the only definitive curable treatment for SCID patients is allogeneic hematopoietic stem cell transplantation (HSCT) from a human leukocyte antigen (HLA)-matched donor<sup>25</sup>. However, finding a HLA-matched donor is rare, and the alternative treatment, haplo-identical HSCT, reduces the survival rate from >80% with an HLA-matched donor to 60–70%<sup>26,27</sup>. Although successful HSCT promotes lymphoid lineage development resulting in a long-term patient survival rate, it is accompanied by a high risk of graft-versus-host disease<sup>28,29</sup>.

An ideal alternative to searching for an HLA-matched donor is to genetically edit the patient's own CD34<sup>+</sup> HSPCs ex-vivo to be subsequently returned to the patient as an autologous HSCT. CD34<sup>+</sup> HSPCs' marked ability to reconstitute the immune system from a small number of cells, makes these cells an attractive platform for gene-therapy applications<sup>30,31</sup>. To that end, transgene delivery via lentiviral (LV) or gammaretroviral ( $\gamma$ RV) vectors for ex-vivo editing of patients' CD34<sup>+</sup> HSPCs was previously reported in gene therapy clinical trials<sup>32–36</sup>. However, genome-editing-based treatments using  $\gamma$ RV of Chronic granulomatous disease (CGD)<sup>37</sup>, Wiskott-Aldrich syndrome (WAS)<sup>38,39</sup>, SCID-X1<sup>40</sup>, and Adenosine deaminase (ADA)-SCID<sup>41</sup> resulted in the activation of proto-oncogenes leading, in some patients, to a leukemic transformation<sup>42,43</sup>. Although steps have been taken to improve the safety of these viral vectors, transgene integration into tumor-suppressor loci has been observed and incomplete phenotypic correction, toxicity, dysregulated hematopoiesis, and insertional mutagenesis related to the semi-random integration and constitutive expression of the transgene into the genome remain major safety concerns<sup>44,45</sup>. Despite these concerns, which are particularly severe for highly controlled and regulated genes such as the *RAG1/2*, LV-based gene therapy for *RAG1* is currently undergoing clinical trials<sup>46</sup>.

For these reasons, transgene delivery via a targeted HDR-mediated genome-editing approach such as the combination of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated nuclease 9 (Cas9) and recombinant adeno-associated virus serotype 6 (rAAV6) could prove particularly beneficial for treating *RAG*-SCIDs<sup>47–49</sup>. Although rAAV6 provides a number of benefits over LV and  $\gamma$ RV vectors, we have shown previously that rAAV6 vectors trigger a toxic DNA damage response (DDR) proportional to the multiplicity of infection (MOI), or the amount of virus used<sup>50</sup>. Thus, for a CRISPR-Cas9/rAAV6 genome-editing strategy to be therapeutically relevant, a delicate balance between maintaining high-quality HDR while reducing viral load as much as possible is required. Therefore, we tested a number of different donor constructs in order to find the optimal rAAV6 donor design for efficient HDR.

Additionally, a highly specific CRISPR-Cas9-based approach is particularly beneficial for disorders where the affected gene is tightly regulated, such as *RAG*-SCIDs<sup>11,12,51</sup>. CRISPR-Cas9/rAAV6 genome editing provides the potential for specific gene integration that maintains

the strict endogenous spatiotemporal gene regulation and expression of the transgene. Modulation of transgene expression levels by including mRNA stability and/or nuclear export cis-acting elements such as polyadenylation (polyA or pA) signals or woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) segments has been widely reported<sup>52</sup>. Polyadenylation at the end of a RNA transcript affects mRNA stability, nuclear export, and translation, thus, playing a crucial role in gene expression. In-vitro studies using rAAV6 vectors have highlighted the effects of different synthetic polyA signals, such as bovine growth hormone (BGH) polyA signal sequences, in boosting transgene expression<sup>53,54</sup>. Similarly, cis-acting post-transcriptional regulatory elements (PREs), such as WPRE, are required for the nuclear export of intronless RNA, and the addition of such PREs has been reported to enhance transgene expression in-vitro<sup>53,55</sup>. Through the use of such sequences to regulate our *RAG2* transgene, we aimed to establish transgene expression that closely resembles that of the endogenous gene.

We previously reported on a proof-of-concept CRISPR-Cas9/rAAV6 *RAG2* gene correction approach<sup>56</sup> via insertion of the corrective transgene into the CRISPR-Cas9-induced cut site adjacent to the *RAG2* start codon. This approach successfully led to the development of T cells with diverse TCR repertoires from *RAG2*-SCID patient-derived CD34<sup>+</sup> HSPCs. Here, we describe a CRISPR-Cas9/rAAV6-mediated genome-editing approach that replaces the entire *RAG2* CDS with a corrective transgene in CD34<sup>+</sup> HSPCs, to maintain the endogenous spatial regulatory elements of the *RAG2* locus. We achieved this by constructing three *RAG2* CDS replacement correction donors where each maintains the 5' endogenous promoter and regulatory elements<sup>10,15,57–59</sup>. However, in addition to the regulatory elements upstream to the *RAG2* gene, at least one silencer sequence has been discovered between the *RAG1* and *RAG2* genes that is counteracted by an anti-silencing element 5' of the *RAG2* promoter<sup>60</sup>. Additionally, there are evolutionarily conserved genes within the *RAG1/2* locus' intronic sequences that are believed to play a role in transcriptional regulation<sup>11,12,61,62</sup>. Thus, to most closely mimic the endogenous expression of *RAG2*, we designed one donor to preserve the 3' UTR and downstream region in addition to the 5' UTR and promoter region. The two additional correction donors were engineered in an effort to increase the expression levels of the transgene by maintaining the *RAG2* 5' UTR endogenous region while introducing synthetic 3' UTRs (WPRE-BGHpA and BGHpA sequences).

Developing a proof-of-concept gene-correction therapy typically requires large amounts of patient samples. However, since untreated SCID patients rarely survive past infancy, neither invasive bone marrow procedures nor drawing large volumes of peripheral blood (PB) are viable options for procuring such samples. In this work, to circumvent this challenge, we used fluorescence-activated cell sorting (FACS) enrichment of healthy donor (HD)-derived CD34<sup>+</sup> HSPCs with engineered genotypes after multiplex HDR (*knock-in/knock-out* [*KI-KO*]) to simulate single-allelic gene-correction therapies for *RAG2*-SCID. Since *RAG2*-SCID is an autosomal recessive disorder, correction of only one mutated allele is sufficient to cure the patient. In this strategy, we mimicked monoallelic correction in SCID-patient cells by *knock-in* (*KI*) of a diverged codon-optimized *RAG2* (*dcoRAG2*) cDNA cassette in place of the endogenous *RAG2* CDS in one allele (thereby preserving regulatory non-coding elements) and by *knock-out* (*KO*) of the second allele with a green fluorescent protein (GFP) gene-disrupting cassette. The *dcoRAG2* cDNA produces a protein identical to wild-type (WT) *RAG2*, while the introduction of wobble changes leads to reduced similarity to the genomic sequence precluding the Cas9 from re-cutting the inserted sequence or from the inserted sequence serving as a homology arm causing premature cessation of HDR. Via cell sorting, we were then able to enrich for cells with the desired *KI-KO*-engineered genotype to model and track their progression into T-cell development. Thus, in this study, we show *RAG2*

gene correction simulation in HD-derived CD34<sup>+</sup> HSPCs by replacement of the entire endogenous CDS.

## Results

### Different HDR strategies: cut-site insertion vs. coding sequence (CDS) replacement and adjusting homology arm length

*RAG2*-SCID is caused by mutations scattered throughout the CDS of the *RAG2* gene<sup>4</sup>. Therefore, a universal correction technique that would suit all *RAG2*-SCID patients requires the delivery of an intact copy of the complete *RAG2* CDS. While *KI* of an intact CDS at the endogenous locus would achieve this, this strategy could interfere with the 3D chromatin architecture and critical endogenous gene regulation by moving regulatory elements further downstream from the transgene. This could potentially disrupt spatial cross-talk between functional elements upstream and downstream of the *RAG2* gene, such as promoter and/or enhancer sequences<sup>12,14,17,18</sup> (Supplementary Fig. 2A–D). Hence, we hypothesized that a donor DNA with a left homology arm (LHA) upstream adjacent to the cut site, and a right homology arm (RHA) distanced from the cut site, downstream of the *RAG2* stop codon would ensure preserving *RAG2* regulatory elements by replacing the entire *RAG2* CDS. To examine the feasibility of our hypothesis, we produced two rAAV6 vectors that would integrate a GFP expression cassette under the regulation of a spleen focus-forming virus (SFFV) promoter and BGHpA sequence after delivery of a chemically modified *RAG2* sgRNA/Cas9 ribonucleoprotein (RNP) complex via electroporation into CD34<sup>+</sup> HSPCs. In previous studies, we demonstrated this sgRNA's high on-target editing efficiency and accuracy<sup>50,63,64</sup>. The first donor<sup>56</sup>, herein CSI\_GFP-BGHPA\_400×400, uses 400 bp homology arms immediately flanking the Cas9-induced cut site for donor insertion (hereafter termed a cut-site-insertion [CSI] vector), while the second donor, herein CDSR\_GFP-BGHPA\_400×400, uses a 400 bp LHA spanning the immediate sequence upstream to the Cas9 cut site and a 400 bp RHA spanning the immediate sequence downstream to the *RAG2* stop codon, to replace the entire *RAG2* CDS with the DNA donor (herein a CDS-replacement [CDSR] vector) (Fig. 1A, Supplementary Table 1, and Supplementary Data 1). Two days post-editing, we analyzed the frequencies of GFP<sup>+</sup> cells via flow cytometry to determine the HDR efficiencies of the different donors. Each rAAV6 donor was gated based on its respective rAAV only (RNP) sample to minimize the number of cells that are included in the positive population that are expressing the reporter in an episomal manner without HDR-mediated integration of the cassette into the genome (GFP<sup>low</sup> cells) (see Supplementary Note 1 and Supplementary Fig. 1 for a description of the gating strategy). We found that the HDR efficiency of the CSI\_GFP-BGHPA\_400×400 vector was significantly higher than that of the CDSR\_GFP-BGHPA\_400×400 vector (21.8% and 9.1%, respectively) (Fig. 1B and Supplementary Fig. 2E). Attempting to improve the HDR efficiency of the CDSR technique, we designed an additional two rAAV6 donors with RHAs extended from 400 bp to 800 bp and 1,600 bp spanning the immediate region downstream to the *RAG2* stop codon (herein CDSR\_GFP-BGHPA\_400×800 and CDSR\_GFP-BGHPA\_400×1600, respectively [Fig. 1A, Supplementary Table 1, and Supplementary Data 1]). While elongation to 800 bp produced significantly higher HDR efficiency than the CDSR\_GFP-BGHPA\_400×400 donor (14.8%), only after elongation to 1,600 bp, did we observe HDR efficiency comparable to that of the CSI\_GFP-BGHPA\_400×400 donor (25.2%). (Fig. 1B and Supplementary Fig. 2E). Using a uniform pair of primers and probe for droplet digital PCR (ddPCR), we confirmed that the HDR efficiencies as determined by flow cytometry were accurate and locus-specific (Fig. 1C and Supplementary Fig. 2F). To validate that our CDSR strategy is broadly applicable and not specific only to the *RAG2* locus, we designed a set of rAAV6 donors to introduce a GFP expression cassette into the *RAG1* locus (Supplementary Fig. 3A, Supplementary Table 2, and Supplementary Data 1). Similar to *RAG2*, we used a highly specific sgRNA that targeted just downstream from the

*RAG1* ATG start codon. Since *RAG1* CDS is longer than that of *RAG2*, the CDSR method here replaced 3,112 bp as opposed to only 1,541 bp at the *RAG2* locus. While we were able to achieve highly efficient HDR at the *RAG1* locus as well, we found that longer homology arms were required to do so (Supplementary Fig. 3B–E).

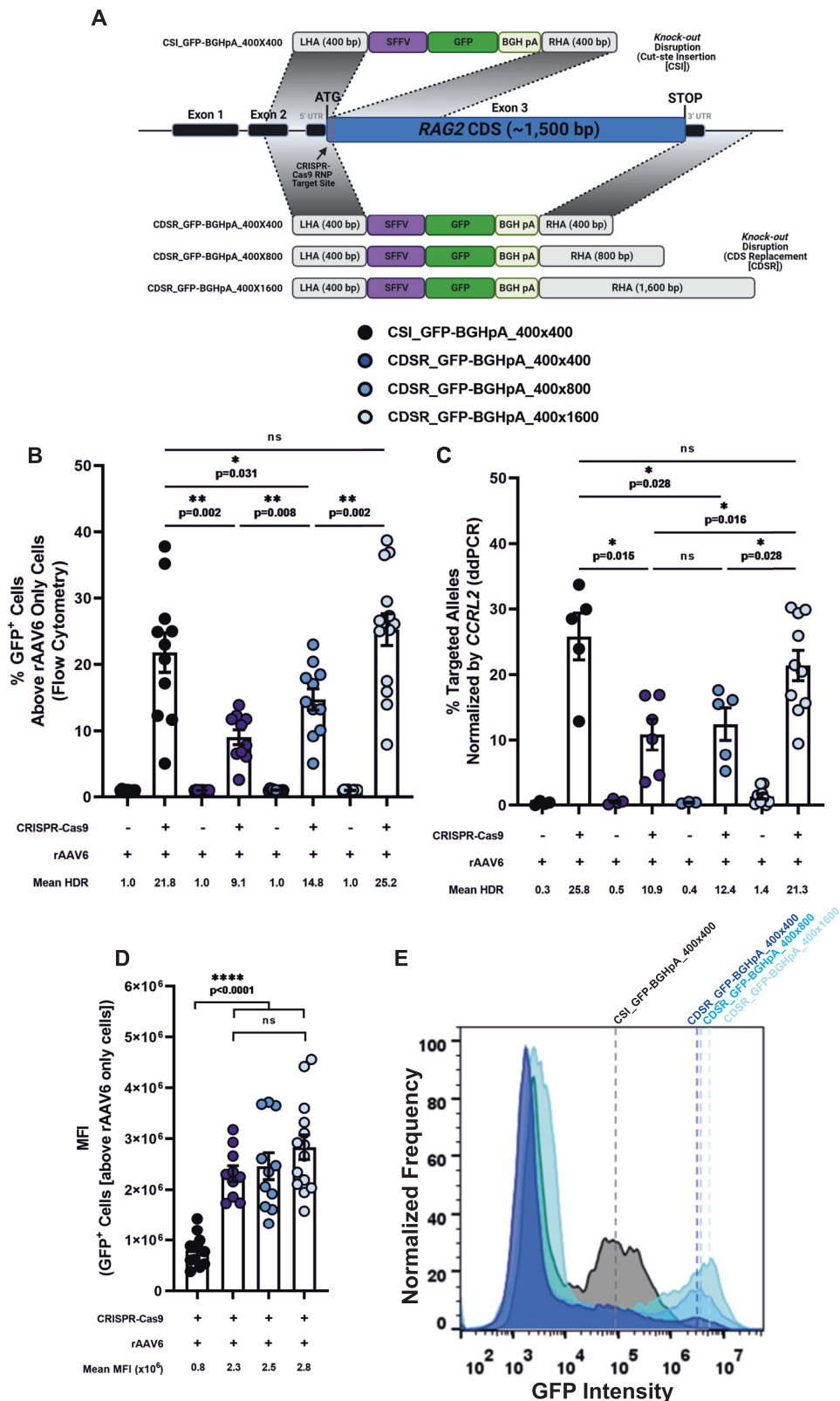
Interestingly, via flow cytometry, we observed a significantly higher mean fluorescence intensity (MFI) of GFP<sup>high</sup> cells after integration of the CDSR vectors compared to the CSI vectors (Fig. 1D, E and Supplementary Fig. 2E and Supplementary Note 1). This difference highlights that different integration strategies have unique effects and can lead to distinctive conformational changes on the genomic locus and impact subsequent transgene expression.

### Synthetic polyA sequences and/or cis-acting PREs affect transgene expression

To modulate transgene expression further, we aimed to test the impact of synthetic 3' regulatory elements on transgene expression. Thus, we designed two CDSR vectors (herein CDSR\_GFP-WPRE-BGHPA\_400×1600 and CDSR\_GFP-NoBGHPA\_400×1600 [Fig. 2A, Supplementary Table 1, and Supplementary Data 1]) with a homology arm pattern similar to that of the CDSR\_GFP-BGHPA\_400×1600 donor. However, whereas the CDSR\_GFP-BGHPA\_400×1600 vector contained a BGHPA sequence alone and the CDSR\_GFP-WPRE-BGHPA\_400×1600 contained a WPRE-BGHPA sequence, the CDSR\_GFP-NoBGHPA\_400×1600 vector lacked both regulatory elements, thus allowing GFP expression to be controlled by the endogenous *RAG2* 3' UTR. While the CDSR\_GFP-BGHPA\_400×1600 and CDSR\_GFP-WPRE-BGHPA\_400×1600 donors produced comparable HDR efficiencies, the CDSR\_GFP-NoBGHPA\_400×1600 donor induced lower HDR as observed by flow cytometry and confirmed by ddPCR (Fig. 2B, C and Supplementary Fig. 4). Interestingly, the three donors produced significantly different MFI levels, with CDSR\_GFP-BGHPA\_400×1600 ( $2.8 \times 10^6$ ) being the highest followed by CDSR\_GFP-WPRE-BGHPA\_400×1600 and CDSR\_GFP-NoBGHPA\_400×1600 ( $1.6 \times 10^6$  and  $0.3 \times 10^6$ , respectively) (Fig. 2D, E), highlighting the strength of the synthetic 3' UTRs in modulating expression patterns (see Supplementary Note 1 and Supplementary Fig. 1 for a description of the gating strategy).

### KI-KO genotype engineering in HD-derived HSPCs using two-part enrichment strategy

Since *RAG2* gene regulation is critical, we aimed to fine-tune our previously published *RAG2*-correction strategy<sup>56</sup> by using the CDSR method. We hypothesized that replacing the entire CDS would allow for transgene expression to be driven by the *RAG2* endogenous promoter and 3' UTR, thus enabling the transgenic dco*RAG2* cDNA expression patterns to most similarly resemble that of endogenous *RAG2*. Additionally, by replacing the entire CDS (~1.5 kb), as opposed to pushing the sequence ~4 kb downstream in the case of HDR via insertion, we are able to more closely maintain the proximity of the *RAG* genes, thus potentially conserving the ability to form the chromatin hub super-enhancer necessary for proper expression. We constructed a CDSR correction donor (herein CDSR\_Corr\_Endo3'UTR [Fig. 3A, Supplementary Table 1, and Supplementary Data 1]) with a 400×800 bp homology arm pattern for *KI* of the dco*RAG2* cDNA. To track the expression of dco*RAG2* cDNA and enrich for cells with successful integration, the dco*RAG2* stop codon was eliminated and replaced with a T2A self-cleaving peptide sequence followed by a truncated nerve growth factor receptor (tNGFR) reporter gene, producing in-frame transcription of the two sequences (dco*RAG2* cDNA and tNGFR). Following the translation of the fusion protein, the T2A self-cleaves producing two proteins (*RAG2* and tNGFR) at a 1:1 ratio in the cell. The use of tNGFR is particularly advantageous since it enables tracking and enrichment of corrected cells and has been approved for clinical applications<sup>65</sup>. Additionally, we constructed two donors with synthetic 3' UTRs following the tNGFR, one with WPRE-BGHPA



sequences and one with only the BGHpA sequence each with a 400x800 bp homology arm pattern (herein CDSR\_Corr\_WPRE-BGHpA and CDSR\_Corr\_BGHpA, respectively [Fig. 3A, Supplementary Table 1, and Supplementary Data 1]). While we observed the highest HDR for CDSR donors with a RHA of 1,600 bp (Fig. 1B, C), we could not design correction donors with a 400x1,600 bp homology arm pattern due to the limited carrying capacity (~4.8 kb) of rAAV6 vectors<sup>66</sup>. For

comparative purposes, we utilized our previously published CSI *Ki* donor which contained *dcoRAG2* cDNA followed by the *RAG2* endogenous 3' UTR sequence along with a tNGFR reporter gene cassette under the regulation of a constitutive phosphoglycerokinase (PGK) promoter and BGHpA sequence between 400 bp homology arms (herein CSI\_Corr)<sup>56</sup> (Supplementary Fig. 5A, Supplementary Table 1, and Supplementary Data 1). We tested these four correction donors

**Fig. 1 | Different strategies for HDR: cut-site insertion vs. CDS replacement and adjusting homology arms.** **A** Schematic representation of *RAG2* KO disruption donors containing a GFP reporter gene cassette under the control of an SFFV promoter and BGHpA sequence. Successful HDR of the CSI\_GFP-BGHPA\_400x400 donors results in the integration of the reporter gene approximately 43 bp downstream from the *RAG2* ATG start codon. Successful HDR of the three CDSR donors results in the replacement of the entire endogenous *RAG2* CDS with the reporter gene cassette. (See Table S1 for a more in-depth description of the donors and Supplementary Data 1 for the exact sequences). Figure was created with BioRender.com. **B** HDR frequencies analyzed by flow cytometry. See Supplementary Note 1 for a description of the gating strategy. CSI\_GFP-BGHPA\_400x400 ( $N=11$ ), CDSR\_GFP-BGHPA\_400x400 ( $N=10$ ), CDSR\_GFP-BGHPA\_400x800 ( $N=11$ ), CDSR\_GFP-BGHPA\_400x1600 ( $N=14$ ). Data are represented as mean  $\pm$  SEM. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$  (Mann-Whitney one-sided

test). **C** Site-specific HDR efficiencies at the *RAG2* locus measured by ddPCR and normalized by targeted *CCRL2* alleles. CSI\_GFP-BGHPA\_400x400 (rAAV6 only:  $N=4$ ; CRISPR + AAV:  $N=5$ ), CDSR\_GFP-BGHPA\_400x400 (rAAV6 only:  $N=4$ ; CRISPR + AAV:  $N=6$ ), CDSR\_GFP-BGHPA\_400x800 (rAAV6 only:  $N=3$ ; CRISPR + AAV:  $N=5$ ), CDSR\_GFP-BGHPA\_400x1600 (rAAV6 only:  $N=9$ ; CRISPR + AAV:  $N=10$ ). Data are represented as mean  $\pm$  SEM. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$  (Mann-Whitney one-sided test). **D** MFI values of HDR<sup>+</sup> cells analyzed by flow cytometry. See Supplementary Note 1 for a description of the gating strategy. CSI\_GFP-BGHPA\_400x400 ( $N=11$ ), CDSR\_GFP-BGHPA\_400x400 ( $N=10$ ), CDSR\_GFP-BGHPA\_400x800 ( $N=11$ ), CDSR\_GFP-BGHPA\_400x1600 ( $N=14$ ). Data are represented as mean  $\pm$  SEM. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$  (Mann-Whitney one-sided test). Source data are provided as a Source Data file. **E** Representative histograms of the MFI induced by the four KO donors.

individually and observed highly effective locus-specific HDR for them all, determined by ddPCR (Supplementary Fig. 5B, C). For confirmation that the integration of the donors occurred as expected, we conducted an ‘in-out’ PCR with one primer located on the tNGFR sequence and one primer downstream to the RHA (Supplementary Fig. 5D). Indeed, the observed amplified bands were consistent with the expected integration patterns and corroborated effective HDR (Supplementary Fig. 5E).

While our correction strategy relies on integrating the rAAV6 donor into the Cas9-induced break site via the homology recombination process, it is known that AAV donor vectors can integrate into random sites in the genome, presumed to be spontaneous DSBs, via the NHEJ pathway<sup>67</sup>. Additionally, the DSBs introduced by CRISPR-Cas9 at on- and off-target sites can also incorporate donor sequences in full or only partially, by NHEJ. In order to assess the specificity of the integration of our corrective donors we took advantage of ITR-seq, a highly effective method to detect integration of the rAAV6 donor’s inverted terminal repeats (ITRs) across the genome<sup>68</sup> (ITR-seq adapters and primers can be found in Supplementary Table 3). Since any off-target integration of the donor would occur via the NHEJ repair pathway, the method is capable of detecting donor integration at any site in the genome<sup>68</sup>. We tested our CSI\_Corr and CDSR\_Corr\_Endo3’UTR donors independently and found that while there was incorporation of the ITRs at the on-target site, there was relatively limited integration of ITRs at other sites in the genome (a single off-target for each donor [Supplementary Table 4]).

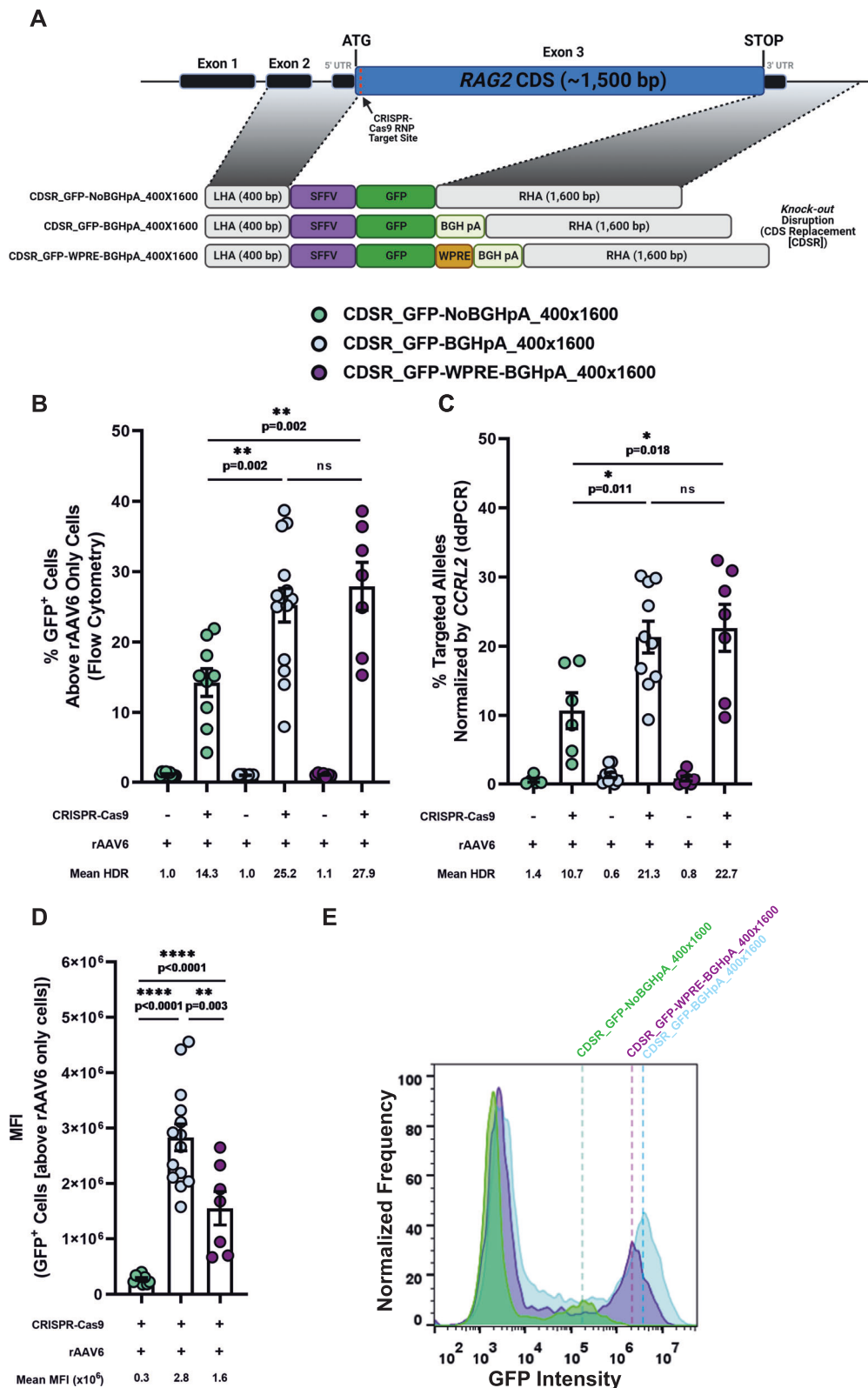
Since the ITR-seq method is not quantitative and only detects sequences with ITR integration, we conducted amplification-free, long-range sequencing via Oxford Nanopore Technologies (ONT) using Cas9-targeted sequencing<sup>69</sup>. This method allows capturing the full scope of events, occurring upon CRISPR-Cas9-based genome editing combined with an rAAV6 donor, at the on-target locus across the cell population, without amplification bias (strategy described in the Methods section). In particular, we were interested in quantitatively assessing the extent of HDR-mediated correction versus NHEJ-based donor integration at the on-target site. Using Cas9-RNP digestion (sgRNA sequences are listed in Supplementary Table 5), we enriched for the on-target locus and analyzed the genome-editing products (Supplementary Fig. 6A, B). We found that across three replicates, the HDR frequencies determined by ONT Ampfree and the HDR frequencies determined by ddPCR (and flow cytometry in the case of CSI\_Corr) were comparable (Supplementary Fig. 6C–I). Individual ONT HDR reads were determined and validated by alignment to a reference sequence displaying ideal HDR integration (Supplementary Fig. 6E, F and Supplementary Data 2). Additionally, NHEJ-based insertions to the cut site and partial NHEJ were kept below 5% and 9%, respectively for the CDSR\_Corr\_Endo3’UTR donor, and below 8% and 4%, respectively for the CSI\_Corr donor, levels that are broadly comparable to prior reports<sup>70–72</sup> (Supplementary Fig. 6C–I and Supplementary Fig. 7). Lastly, we detected premature cessation of HDR when editing with the

CSI\_Corr donor (4.2%), due to the presence of the 3’ UTR sequence in the donor (Supplementary Fig. 6D and Supplementary Fig. 7). In these cases, the non-diverged 3’ UTR sequence in the CSI\_Corr donor acted as a 3’ homology arm with the identical endogenous 3’ UTR sequence and led to incomplete HDR.

We aimed to utilize the *KI-KO* strategy to engineer genotypes via multiplex HDR in HD-derived CD34<sup>+</sup> HSPCs to simulate the therapeutic outcome of *RAG2*-SCID single-allelic correction following a gene-editing-based treatment. This strategy has two main advantages over more extensive editing methodologies: 1) In contrast to the use of induced pluripotent stem cells (iPSCs)<sup>73–76</sup>, HD-derived CD34<sup>+</sup> HSPCs are biologically authentic since they are the same cells used in HSCT;<sup>77</sup> and 2) Lengthy culturing protocols are inadequate since CD34<sup>+</sup> HSPCs lose their regenerative ability as well as their engraftment potential after prolonged culturing<sup>78</sup>. Thus, we chose to apply our *KI-KO* strategy in HD-derived CD34<sup>+</sup> HSPCs by utilizing multiplex HDR to obtain a cell population with one allele targeted with one of the four aforementioned correction donors and the other allele with a *KO* template (Fig. 1A [CDSR\_GFP-BGHPA\_400x800 was paired with the three CDSR correction donors and CSI\_GFP-BGHPA\_400x400 was paired with the CSI\_Corr]). We utilized our *RAG2*-SCID disease model, reported on in Iancu et al., in order to compare to our correction simulation results<sup>56</sup>.

For the CSI\_Corr donor, enrichment of *KI-KO* CD34<sup>+</sup> HSPCs was achieved by sorting for biallelic double-positive tNGFR<sup>+</sup>/GFP<sup>+</sup> expression two days post-electroporation (herein day 0) and immediately seeding the cells into the in-vitro T-cell differentiation (IVTD) system (Supplementary Fig. 8A, B). However, with CDSR correction donors, since tNGFR expression is under the regulation of the endogenous *RAG2* promoter (and the *RAG2* expression window occurs later in the T-cell developmental process), there is no expression of tNGFR on day 0. Therefore, a unique enrichment strategy was required to isolate *KI-KO* cells. On day 0, the CDSR donors were sorted only for *KO* GFP expression, and the GFP<sup>+</sup> cells were immediately seeded into the IVTD system. On day 14 of IVTD, when *RAG2* is highly expressed (thus, tNGFR is also expressed [Supplementary Fig. 8C]) tNGFR<sup>+</sup> cells were sorted for and seeded back into the IVTD system (Fig. 3B, Supplementary Fig. 8D, and Supplementary Fig. 9). Additionally, all samples were sorted for CD7 expression to enrich for cells that have begun to differentiate, namely, only cells that were CD7<sup>+</sup> were subjected to days 14–28 of IVTD (Supplementary Fig. 9). ddPCR was performed on genomic DNA from the *KI-KO* populations to confirm that the two-step enrichment method indeed led to the enrichment of a cell population with ~100% edited alleles. Indeed, in all four correction donor multiplex HDR combinations, ~100% of targeted alleles were found to be positive (Fig. 3C and Supplementary Fig. 8E).

Additionally, since the CDSR donors produce a fusion protein separated by a self-cleaving T2A sequence resulting in a 1:1 ratio between transgenic *RAG2* and tNGFR, tNGFR MFI measurement was used as a proxy measurement for transgenic dco*RAG2* expression levels. As expected, we observed that the MFI for CDSR\_Corr\_WPRE-



BGHpA was 2x that of CDSR\_Corr\_Endo3'UTR and 1.4x greater than CDSR\_Corr\_BGHpA on day 28 of IVTD with a similar trend on day 14 (Fig. 3D, E) indicating higher expression of the *dcoRAG2*.

### *KI-KO* HSPCs Produce CD3<sup>+</sup>TCR $\gamma\delta$ <sup>+</sup> and CD3<sup>+</sup>TCR $\alpha\beta$ <sup>+</sup> T Cells in the IVTD System

With a robust method to isolate cells with the *KI-KO* genotype, we aimed to validate that specifically the expression of *dcoRAG2* enabled

the *KI-KO* cells to differentiate into CD3<sup>+</sup> T cells with diverse TCR repertoires, to present a proof-of-concept for gene correction. Quantitative real-time PCR (qRT-PCR) using transcript-specific primer pairs (Supplementary Fig. 10A) revealed that the expression of endogenous *RAG2* was ostensibly eliminated in the *KI-KO* populations while robust *dcoRAG2* cDNA expression was found exclusively in the *KI-KO* engineered cells (Fig. 4A and Supplementary Fig. 10B). Additionally, when we directly compared the total *RAG2* mRNA levels between all groups,

**Fig. 2 | Effect of synthetic polyA and/or cis-acting PREs on transgene expression.** **A** Schematic representation of *RAG2* KO disruption donors containing a GFP reporter gene cassette under the control of an SFFV promoter to test the effect of different formulations of the 3' UTR. (Top to bottom) CDSR\_GFP-NoBGHpA\_400x1600, CDSR\_GFP-BGHpA\_400x1600 (data presented in Fig. 1A), and CDSR\_GFP-WPRE-BGHpA\_400x1600. Successful HDR of the three donors results in the replacement of the entire endogenous *RAG2* CDS with the reporter gene cassette. (See Supplementary Table 1 for a more in-depth description of the donors and Supplementary Data 1 for the exact sequences). Figure was created with BioRender.com. **B** HDR frequencies analyzed by flow cytometry. See Supplementary Note 1 for a description of the gating strategy. CDSR\_GFP-NoBGHpA\_400x1600 ( $N=9$ ), CDSR\_GFP-BGHpA\_400x1600 ( $N=14$ ), and CDSR\_GFP-WPRE-BGHpA\_400x1600 ( $N=7$ ). Data are represented as mean  $\pm$  SEM. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$  (Mann-Whitney one-sided test). **C** Site-

specific HDR efficiencies at the *RAG2* locus measured by ddPCR and normalized by targeted *CCR2L2* alleles. CDSR\_GFP-NoBGHpA\_400x1600 (IRAAV6 only:  $N=5$ ; CRISPR + AAV:  $N=6$ ), CDSR\_GFP-BGHpA\_400x1600 (IRAAV6 only:  $N=9$ ; CRISPR + AAV:  $N=10$ ), and CDSR\_GFP-WPRE-BGHpA\_400x1600 (IRAAV6 only:  $N=7$ ; CRISPR + AAV:  $N=7$ ). Data are represented as mean  $\pm$  SEM. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$  (Mann-Whitney one-sided test). **D** MFI values of HDR<sup>+</sup> cells analyzed by flow cytometry. See Supplementary Note 1 for a description of the gating strategy. CDSR\_GFP-NoBGHpA\_400x1600 ( $N=9$ ), CDSR\_GFP-BGHpA\_400x1600 ( $N=14$ ), and CDSR\_GFP-WPRE-BGHpA\_400x1600 ( $N=7$ ). Data are represented as mean  $\pm$  SEM. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$  (Mann-Whitney one-sided test). Source data are provided as a Source Data file. **E** Representative histograms of the MFI induced by the three KO donors.

we found that expression of the dco*RAG2* transgenes does not exceed that of the *Mock* samples indicating that the transcription is still tightly controlled and that the gene is not being overexpressed (Fig. 4B).

Importantly, the expression of the dco*RAG2* cDNA indeed facilitated T-cell development highlighted by the successful differentiation of *RAG2* *KI-KO* cells into CD7<sup>+</sup>, CD5<sup>+</sup>, and CD1a<sup>+</sup> pre-T cells on day 14 and CD3<sup>+</sup> T cells by day 28 (Fig. 4C, D and Supplementary Fig. 10C, D [gating strategy depicted in Supplementary Fig. 11]). Additionally, robust TCR $\gamma\delta$  expression in the CD3<sup>+</sup> population was observed by flow cytometry on day 28 with the CD3<sup>+</sup>TCR $\gamma\delta$ <sup>+</sup> cells presumed to be CD3<sup>+</sup>TCR $\alpha\beta$ <sup>+</sup> T cells (Fig. 4E). Lastly, PCR amplification using primers flanking the V-J regions of the TRG locus highlighted the successful recombination of *KI-KO* cells comparable to that of the *Mock* cells on day 28 (Supplementary Fig. 10E).

### Expression of dco*RAG2* cDNA induces normal TCR repertoire development

Deep-sequencing analysis of TRB and TRG recombination on day 28 revealed diverse V(D)J rearrangement repertoires in the *RAG2* *KI-KO* populations following expression of the dco*RAG2* cDNA (Fig. 5A) with no significant differences in either TRB or TRG clonotypes between the *Mock* and *RAG2* *KI-KO* populations as calculated by Shannon's H and Simpson's 1-D diversity indices (Fig. 5B, C). Lastly, the complementarity determining region 3 (CDR3) lengths frequency distribution was comparable in all *RAG2* *KI-KO* and *Mock* populations for both the TRB and TRG sequencing (Supplementary Fig. 12). CDR3 is the region of the TCR responsible for recognizing processed antigen peptides and its sequence and length varies from one clone to another. Thus, sequencing the CDR3 regions of a cell population is used as a measurement of TCR diversity<sup>79</sup>. Together, these data indicate that *KI* and expression of the dco*RAG2* cDNA promotes successful V(D)J recombination, subsequent differentiation into CD3<sup>+</sup>TCR $\alpha\beta$ <sup>+</sup> and CD3<sup>+</sup>TCR $\gamma\delta$ <sup>+</sup> T cells, and the development of highly diverse TRB and TRG repertoires.

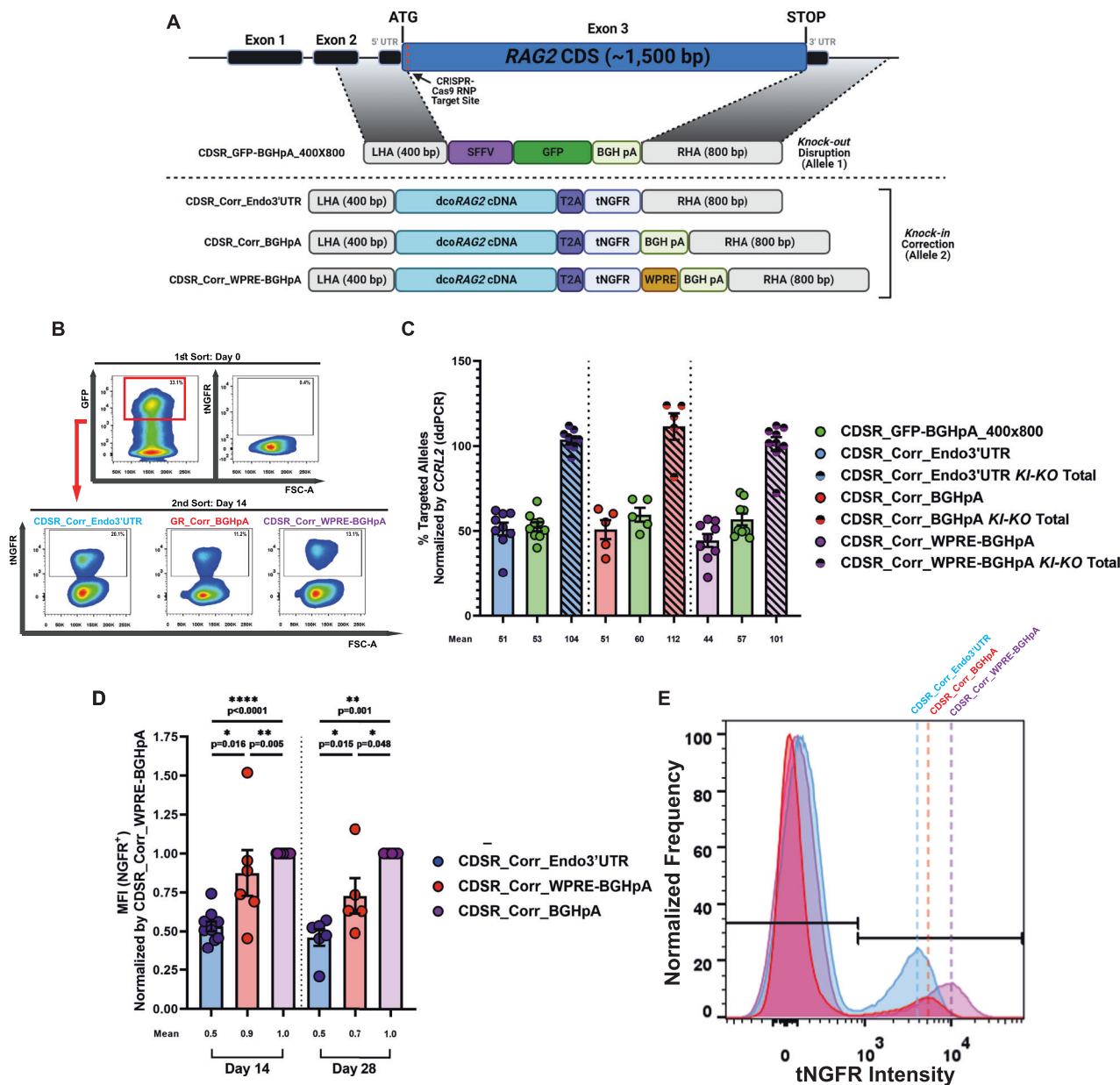
### Discussion

When it comes to highly regulated genes such as *RAG2*, a correction strategy that replaces the entire endogenous CDS may be particularly advantageous, since it provides the ability to maintain the gene's endogenous regulatory elements and safeguard the locus architecture (see Supplementary Fig. 2A–D)<sup>53</sup>. Previous works to correct the *RAG* genes lacked the ability to maintain endogenous regulatory and spatiotemporal elements since the integration of the transgene was either semi-random or via insertion of thousands of bp to the Cas9-induced cut site, thus distancing the *RAG* genes one from the other and potentially altering the genomic locus and hindering optimal gene expression. While the extent to which the CDSR strategy provides an improvement regarding this central issue requires additional follow-up studies, our previously reported corrective donor (referred to in this work as CSI\_Corr) has two

additional shortcomings that our CDSR correction donors aimed to improve upon<sup>56</sup>. Firstly, as outlined in Hubbard et al., the non-diverged 3' UTR sequence of the CSI\_Corr donor can act as a 3' homology arm with the identical endogenous 3' UTR sequence and lead to incomplete or early cessation of HDR<sup>80</sup>. Indeed, when we analyzed the gene-editing products after editing with the CSI\_Corr donor by Ampfree long-read sequencing we identified events where the nondiverged 3' UTR donor sequence acted as a 3' homology arm leading to early cessation of HDR. While Gardner et al. and Pavel-Dinu et al. were able to avoid this possibility by introducing a BGHpA or WPRE-BGHpA sequence in place of the 3' UTR, we believe that this is an inferior solution to relying on the endogenous regulatory sequence<sup>73,81</sup>. Secondly, the incorporation of a complete reporter cassette in the CSI\_Corr donor can have major implications on local chromatin structure and regulation. Chiefly concerning is the presence of a constitutive PGK promoter. The insertion of such an element in a genomic locus like *RAG1/2* that requires such tight regulation is a risk that would be ideal to eliminate. Our CDSR strategy eliminates both the need to incorporate the 3' UTR to the donor (via the CDSR strategy, the expression of the transgene relies on the endogenous genomic 3' UTR region) as well as the need to incorporate a potentially problematic constitutive PGK promoter (via the CDSR strategy, the tNGFR cassette is tied to the dco*RAG2* by separating them with a T2A self-cleaving element). This enables us not only to track *RAG2* expression but to eliminate the need for external promoters.

In our work, we show that the CDSR method is able to induce efficient expression of the dco*RAG2* cDNA without exceeding naturally occurring *RAG2* expression levels found in *Mock* samples (25–50% expression of *RAG2*). Taken together with the efficient T-cell differentiation induced by the corrective transgenes, we describe an approach to simulating genome-editing correction for the treatment of autosomal recessive inborn errors of immunity. While our approach specifically addresses *RAG2*, the methodology can be applied to any tightly controlled gene.

Interestingly, when we designed a CDSR rAAV6 donor by distancing the RHA from the cut site (CDSR\_GFP-BGHpA\_400x400), HDR frequencies were significantly reduced compared to the insertion donor with equivalent homology arm lengths (CSI\_GFP-BGHpA\_400x400). Recently, Cromer et al. showed that elongating the length of the homology arms can lead to higher rates of HDR<sup>53</sup>. Thus, to account for the reduced efficiencies with the CDS replacement donor, we elongated the RHA to 800 bp which induced higher HDR frequencies, however, only once we elongated the RHA to 1,600 bp were we able to completely abrogate the low HDR efficiency observed with 400 bp. We hypothesize that by increasing the length of the distal homology arm, we enhanced the possibility for recognition and subsequent incorporation of the exogenous donor template by the cellular HDR mechanism. Additionally, we observed a significant change in expression of the reporter gene between insertion and CDS



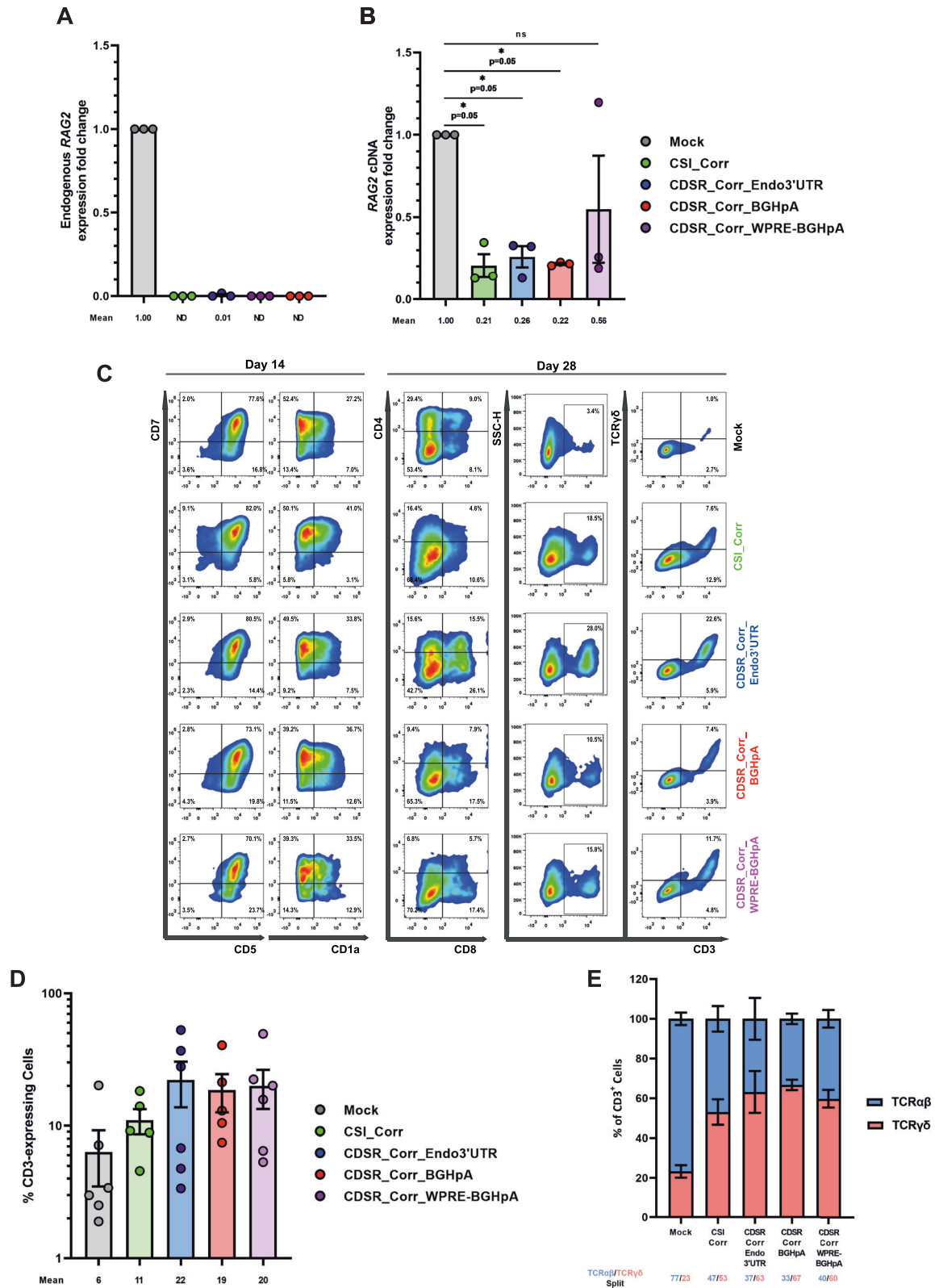
**Fig. 3 | KI-KO simulation of functional gene correction of RAG2 in HD-derived HSPCs using two-part sort strategy.** **A** Schematic representation of RAG2 donors for KI-KO correction simulation. (Top to bottom) CDSR\_GFP-BGHpA\_400x800: RAG2 disruption donor for gene KO (presented in Fig. 1A), CDSR\_Corr\_Endo3'UTR, CDSR\_Corr\_BGHpA, and CDSR\_Corr\_WPRE-BGHpA: RAG2 correction donors for KI of a dcoRAG2 cDNA sequence. Successful HDR of the four donors results in the replacement of the entire endogenous RAG2 CDS. (See Supplementary Table 1 for an in-depth description of the donors and Supplementary Data 1 for the exact sequences). Figure was created with BioRender.com. **B** Two-step FACS enrichment approach for KI-KO multiplexed HDR gene-targeted CD34<sup>+</sup> HSPCs post-CRISPR-Cas9/rAAV6 editing with a combination of either CDSR\_GFP-BGHpA\_400x800 and CDSR\_Corr\_Endo3'UTR, CDSR\_GFP-BGHpA\_400x800 and CDSR\_Corr\_BGHpA, or CDSR\_GFP-BGHpA\_400x800 and CDSR\_Corr\_WPRE-BGHpA. (Top row) Representative FACS plots of the populations two days post-editing (day 0). All three correction groups express GFP yet do not express tNGFR since the RAG2 locus does not undergo transcription until later in the T-cell differentiation process.

Enrichment for GFP<sup>+</sup> cells is conducted, and cells are seeded into the IVTD system. (Bottom row) Representative FACS plots of the populations after 14 days in IVTD (day 14). All three correction groups express tNGFR at different levels. Enrichment for tNGFR<sup>+</sup> cells produces a homogenous double-positive tNGFR<sup>+</sup>/GFP<sup>+</sup> population indicative of KI-KO allelic integration. These cells are seeded into the IVTD system for another 14 days. **C** Site-specific multiplex HDR efficiencies at the RAG2 locus measured by ddPCR and normalized by targeted CCRL2 alleles. CDSR\_Corr\_Endo3'UTR (N = 9), CDSR\_Corr\_BGHpA (N = 5), and CDSR\_Corr\_WPRE-BGHpA (N = 10). Data are represented as mean ± SEM. **D** tNGFR MFI indicating the level of transgenic RAG2 expression on days 14 and 28 of IVTD normalized to MFI of CDSR\_Corr\_WPRE-BGHpA. CDSR\_Corr\_Endo3'UTR ([day 14: N = 10, day 28: N = 6]), CDSR\_Corr\_BGHpA ([day 14: N = 6, day 28: N = 5]), and CDSR\_Corr\_WPRE-BGHpA ([day 14: N = 11, day 28: N = 6]). Data are represented as mean ± SEM. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001, and \*\*\*\* p < 0.0001 (Mann-Whitney one-sided test). Source data are provided as a Source Data file. **E** Representative histograms of the MFI induced by the three KI correction donors on day 14 of IVTD.

replacement donors, attributable to substantial and unique conformational changes in the edited locus which affect expression consistently over HDR at both the RAG1 and RAG2 loci. Lastly, we found that we were able to further modulate the reporter gene expression via the addition or removal of synthetic polyA sequences. Synthetic 3'

UTRs are important and easily implementable tools, some of which have been already employed in gene therapy applications<sup>46,82</sup>. Although these strategies proved to be effective for improving HDR and/or reporter-gene expression in the RAG2 locus, it is well documented that every genomic locus has its own unique characteristics





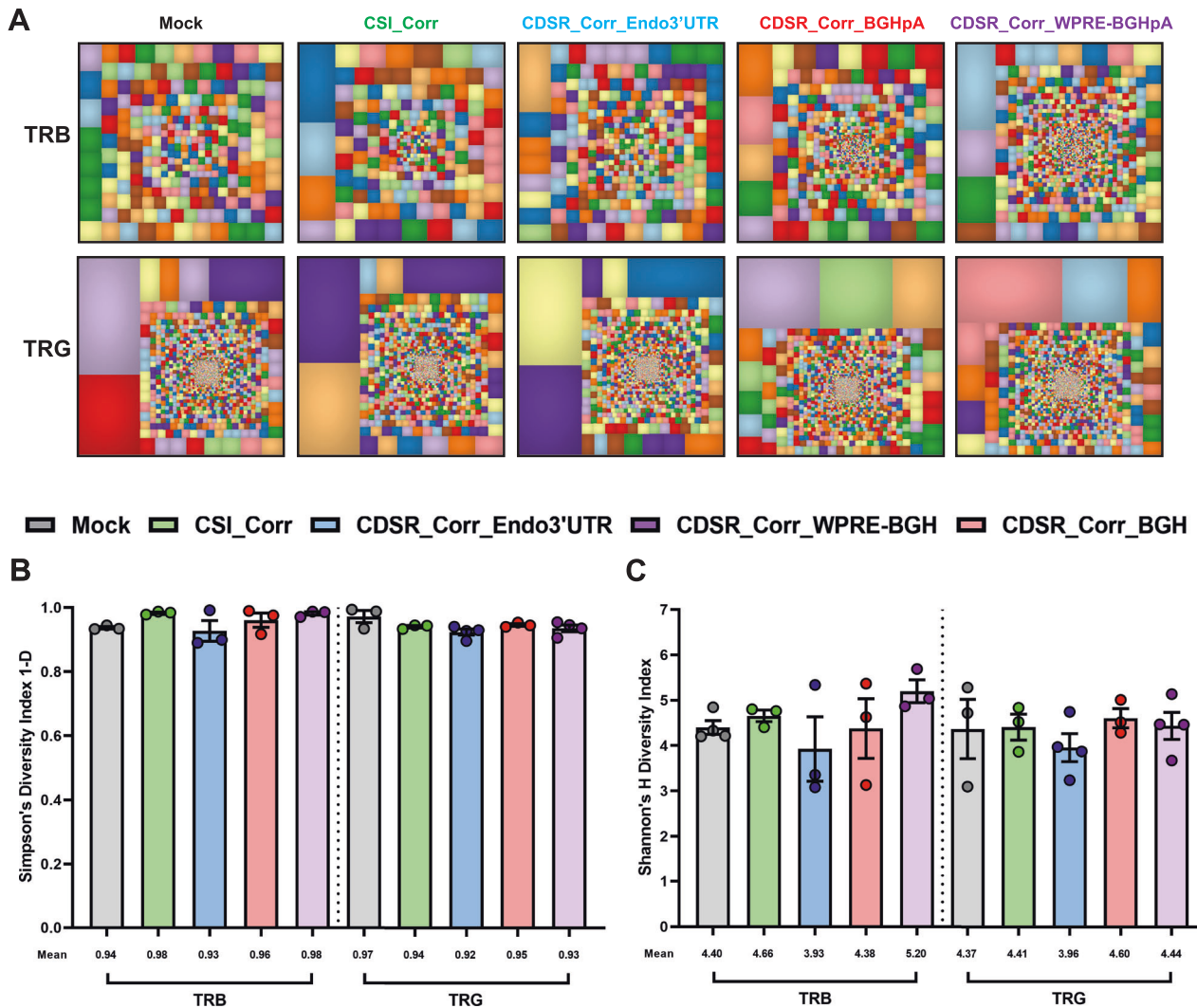
and genomic topology. Thus, we surmise that the exact length and positioning of homology arms to induce efficient HDR in both an integration and CDS replacement manner may be target- and locus-specific.

To establish that the expression of our transgene alone is capable of differentiating CD34<sup>+</sup> HSPCs into CD3-expressing T cells with diverse TRB and TRG repertoires, we utilized a *Kl-KO* approach for each of our CDS replacement correction donors in HD-derived CD34<sup>+</sup>

HSPCs. Since large amounts of SCID-patient samples are scarce, many studies have used iPSCs to develop gene-correction models<sup>73-76</sup>. However, in contrast to iPSCs, HD-derived CD34<sup>+</sup> HSPCs provide a biologically authentic platform since they are the same cells used in HSCT<sup>77</sup>. The use of the *Kl-KO* strategy outlined in Iancu et al. to mimic single-allelic correction of SCID-patient cells provides that all T-cell differentiative progress is due solely to the expression of the transgene and not due to the presence of endogenous *RAG2*<sup>56</sup>. *Kl* of the transgene

**Fig. 4 | IVTD of *KI-KO* correction simulation cells produces CD3<sup>+</sup>TCR $\gamma$  $\delta$ <sup>+</sup> and CD3<sup>+</sup>TCR $\alpha$  $\beta$ <sup>+</sup> T cells.** **A** qRT-PCR quantification of endogenous *RAG2* gene expression in the *RAG2 KI-KO* cells compared to *Mock* cells on day 28 of IVTD (using green primer pair depicted in Supplementary Fig. 10A). Expression fold change is plotted relative to *Mock* and samples with no expression detected are plotted as not determined (ND). (*N* = 3 biologically independent samples). Data are represented as mean  $\pm$  SEM. **B** qRT-PCR quantification of total *RAG2* expression in the *RAG2 KI-KO* cells compared to *Mock* cells on day 28 of IVTD (using red primer pair depicted in Supplementary Fig. 10A). Expression fold change is plotted relative to *Mock*. (*N* = 3 biologically independent samples). Data are represented as mean  $\pm$  SEM. \* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001, and \*\*\*\* *p* < 0.0001 (Mann-Whitney one-sided test). **C** Flow

cytometry analysis of the T-cell developmental progression on days 14 and 28 of IVTD of *Mock* and *KI-KO* populations. Cells were stained for CD7, CD5, and CD1a expression on day 14 of IVTD and for CD4, CD8, CD3, and TCR $\gamma$  $\delta$  expression on day 28 of IVTD. Gating was determined by FMO + isotype controls. **D** Summary of CD3 expression by *Mock* and *KI-KO* populations on day 28 of IVTD. *Mock* (*N* = 6), CSI\_Corr (*N* = 5), CDSR\_Corr\_Endo3'UTR (*N* = 6), CDSR\_Corr\_BGHpA (*N* = 5), and CDSR\_Corr\_WPRE-BGHpA (*N* = 6). Data are represented as mean  $\pm$  SEM. **E** Division between TCR $\alpha$  $\beta$ - and TCR $\gamma$  $\delta$ -expressing CD3<sup>+</sup> cells for *Mock* and *KI-KO* populations on day 28 of IVTD. *Mock* (*N* = 6), CSI\_Corr (*N* = 5), CDSR\_Corr\_Endo3'UTR (*N* = 6), CDSR\_Corr\_BGHpA (*N* = 5), and CDSR\_Corr\_WPRE-BGHpA (*N* = 6). Data are represented as mean  $\pm$  SEM. Source data are provided as a Source Data file.



**Fig. 5 | Expression of *dcoRAG2* cDNA induces normal TCR $\alpha$  $\beta$  and TCR $\gamma$  $\delta$  repertoire development.** **A** Representative tree map depiction of the clonal complexity of the TRB and TRG deep-sequencing repertoires of differentiated T cells from *Mock* and *KI-KO* populations. Within each treemap, each square represents a unique V-J recombination product, and the size of each square represents the clone's relative frequency. **B**, **C** Simpson's 1-D and Shannon's H

diversity indices of TRB and TRG repertoires on day 28 of *Mock* and *KI-KO* populations. *Mock* ([TRB: *N* = 3 and TRG: *N* = 3]), CSI\_Corr ([TRB: *N* = 3 and TRG: *N* = 3]), CDSR\_Corr\_Endo3'UTR ([TRB: *N* = 3 and TRG: *N* = 4]), CDSR\_Corr\_BGHpA ([TRB: *N* = 3 and TRG: *N* = 3]), and CDSR\_Corr\_WPRE-BGHpA ([TRB: *N* = 3 and TRG: *N* = 4]). Data are represented as mean  $\pm$  SEM. Source data are provided as a Source Data file.

in one allele and reliance on Cas9-induced insertions or deletions (INDELS) for *KO* of the second allele would not be sufficient since the Cas9 does not induce *KO* INDELS in 100% of the relevant alleles (50% of the total alleles), and there would be no way to isolate the *KI-KO* cells in this case. Thus, using multiplex HDR with two distinct reporter genes to indicate the *KI-KO* genotype and FACS enrichment of the double-positive population is the most effective method for achieving ~100%

*KI-KO* enrichment in HD-derived CD34<sup>+</sup> HSPCs. Since our correction donors rely on the endogenous *RAG2* promoter for expression which only begins seven days into IVTD (see Supplementary Fig. 8C), we developed a two-step enrichment process over the course of T-cell differentiation which provided a robust platform for testing our *dcoRAG2* correction donors (Supplementary Fig. 9). Additionally, due to the rapid expansion of cells over the first 14 days of IVTD,

our two-step enrichment strategy enabled us to both use fewer cells and to reduce the MOI from 25,000 viral genomes (VG)/cell (used in Figs. 1 and 2) to 12,500 VG/cell while still maintaining requisite HDR. This is particularly beneficial since our group previously showed, in Allen et al., that rAAV6 vectors can trigger a potentially toxic DDR after entering cells, proportional to the MOI used<sup>50</sup>. Thus, reduction of the rAAV6 toxicity is required in order for a rAAV6-based method to be implemented as a clinical therapy for the purpose of gene correction. Follow-up studies to test different methods of improving HDR to allow for further reduction of the MOI, including inhibiting the NHEJ repair pathway via molecules such as i53 and/or DNA-PK inhibitors<sup>83–85</sup>, are currently underway.

Utilizing the *KI-KO* strategy, we were able to track the different correction donors individually and pinpoint the effect that their expression patterns had on T-cell development. We observed higher *dcoRAG2* mRNA levels in the CDS replacement correction donors than in the insertion correction samples, highlighting the effect of different gene-editing and transgene-integration strategies on the resulting gene expression. Notably, even between CDS replacement donors, we observed a significantly higher level of *dcoRAG2* mRNA when the transgene expression was under the 3' regulation of the WPRE-BGHpA sequence than the expression level under the endogenous 3' UTR. Additionally, the *RAG2* mRNA levels produced by the CDS replacement donors were observed to be 25–50% that of the *Mock* samples indicating that overexpression of *RAG2* transgene is not occurring. We expect that the difference between mRNA levels in *Mock* and *KI-KO* samples is in part a function of only one active allele in the correction simulation samples compared to two active alleles in the *Mock* samples. Since parents of *RAG2*-SCID patients (carriers of *RAG* mutations [“naturally-occurring” *KI-KO*]) present clinically as normal and together with the fact that our *KI-KO* samples are able to differentiate effectively, we conclude that the mRNA levels observed are sufficient. Lastly, since the *dcoRAG2* and tNGFR are expressed from the CDS replacement donors as a fusion protein, when the T2A peptide that separates the two proteins self-cleaves, the result is a 1:1 ratio of *RAG2*:tNGFR. Thus, the MFI of tNGFR can serve as a proxy measurement for the level of translation of *RAG2* in the cell. Indeed, higher MFI was observed for the CDS replacement donor under the 3' regulation of the WPRE-BGHpA sequence, consistent with the difference identified on the mRNA level. Together, these data indicate that the synthetic 3' UTRs have a significant effect on the mRNA stability and/or nuclear export efficiency leading to greater translation into *RAG2*. Lastly, all three CDS replacement *RAG2* correction donors promoted successful V(D)J recombination and subsequent differentiation into CD3<sup>+</sup>TCR $\alpha\beta$ <sup>+</sup> and CD3<sup>+</sup>TCR $\gamma\delta$ <sup>+</sup> T cells and developed highly diverse TRB and TRG repertoires. Although the CDS replacement correction donors with synthetic 3' UTRs may be efficient for demonstrating robust transgene expression and successful T-cell development in our IVTD system, they retain the risk of leading to aberrant expression patterns in patient cells. While more extensive studies are ongoing in our lab to assess this possibility, we believe that the CDSR\_Corr\_Endo3'UTR holds the most promise due to its marked ability to induce CDS replacement while conserving the regulatory elements on both the 5' and 3' ends of the *RAG2* gene. Since the 3D genomic architecture of the *RAG* locus is critical for proper gene expression, we expect that follow-up studies will corroborate that this strategy is the optimal one.

In summary, the aforementioned results describe a genome-editing strategy that replaces the entire *RAG2* CDS (1,584 bp [in contrast to the CSI\_Corr donor that leads to an introduction of an additional 3,815 bp at the Cas9 cut site]), thus treating any and all possible mutations in the gene while introducing a transgene that maintains endogenous regulatory and spatiotemporal elements. While we only present data for the two *RAG* genes (maximum replacement of 3,112 bp), we believe that this replacement method can be applied to other genes with a singular coding exon. Additionally, this strategy can

be employed for exon replacement where known mutations are localized on a single exon. Gray et al. showed that for a gene with multiple coding exons, the elimination of critical introns in the donor template led to a significant decrease in transgene expression<sup>86</sup>. Thus, selectively replacing an entire exon while retaining all critical introns could be of great importance. We believe that the CRISPR-Cas9/rAAV6-based CDS replacement strategy for a proof-of-concept gene therapy, which is site-specific and in a clinical setting would use a patient's own cells, can alleviate a number of current barriers to treatment for not only *RAG2*-SCID but other monogenic diseases of the blood and immune system as well. First, autologous gene therapy would eliminate the need for finding an HLA-matched donor and would reduce the risk of graft rejection. Second, using a CRISPR-Cas9/rAAV6-based approach avoids the risk of dysregulated hematopoiesis, incomplete phenotypic correction, and insertional mutagenesis associated with the semi-random integration of a  $\gamma$ RV- or LV-based approach. Third, since the 3D genomic architecture of the *RAG1/2* locus is critical for proper gene expression<sup>12,14,17,18</sup> we were able to avoid inserting kilobases of new DNA between the two genes by using a CDS replacement strategy. This allows for transgene integration while maintaining important regulatory and spatiotemporal elements which can be crucial, especially in genes like *RAG2* that are expressed in a very limited and tightly controlled window. While follow-up studies are undergoing to elucidate the potential benefits of this third point even further, here we show that expression patterns of the corrective transgene are both sufficient to induce quality T-cell differentiation while not exceeding endogenously occurring levels in the *Mock* samples. For these reasons, we believe that the adoption of such a strategy can assist in reducing the risks associated with gene therapy and lead to safer and more accurate applications.

## Methods

### Ethical statement

Our research complies with all relevant ethical regulations. Human cord blood (CB)-derived CD34<sup>+</sup> HSPCs, irrespective of sex, were obtained from Sheba Medical Center CB bank under Institutional Review Board-approved protocols (Approval 3500-16-SMC). Donations of CB are collected from the obstetric delivery department after informed consent is received allowing for CB units that are not suitable for banking to be used for research purposes. No compensation is given to the donors upon consent.

### Cells and cell-culture conditions

CD34<sup>+</sup> HSPCs were isolated via magnetic bead separation (Miltenyi Biotec). CD34<sup>+</sup> HSPCs were cultured in SFEM II enriched with 100 ng/ml Flt3-Ligand, 100 ng/ml TPO, 100 ng/ml SCF, 0.035 mg/ $\mu$ l UM171, 0.75 mg/ $\mu$ l SRI (Stemcell Technologies), 20 unit/ml penicillin, and 20 mg/ml streptomycin (Biological Industries, Beit Haemek, Israel) at 37 °C, 5% CO<sub>2</sub>, and 5% O<sub>2</sub>. Each repeat in this paper was performed on biologically unique and independent CD34<sup>+</sup> HSPCs from different CB donors.

### CRISPR-Cas9 preparation and nucleofection

CRISPR-Cas9 RNP complex preparation and nucleofection were conducted in accordance with the extensive protocol published in Shapiro et al.<sup>87</sup>. A single guide RNA (sgRNA, Alt-R<sup>®</sup> sgRNA, IDT) with end chemical modifications was complexed with Cas9 at a molar ratio of 1:2.5 (Cas9:sgRNA) for 10–20 min at 25 °C to form the RNP complex. The *RAG2* sgRNA variable region sequence is 5'-UGAGAAGCCUGGCU-GAAUUA-3' and the *RAG1* sgRNA variable region sequence is 5'-UUGACUCAGGGUCCACCCA-3'. RNP complexes were added to CD34<sup>+</sup> HSPCs reconstituted in P3 Primary Cell electroporation solution, according to the manufacturer's protocol (Lonza) at a final molar concentration of 4  $\mu$ M. The cell solution was electroporated in the Lonza 4D-Nucleofector using the DZ-100 program.

### rAAV6 DNA donor design and vector production

All rAAV6 vector plasmids were designed and cloned into the pAAV-MCS plasmid containing AAV2-specific ITRs. The pDGM6 plasmid, containing the AAV6 cap genes, AAV2 rep genes, and adenovirus helper genes, was a gift from David Russell (University of Washington). The final rAAV6 vectors were produced by The University of North Carolina (UNC) Vector Core in large-scale rAAV6 batches (UNC Vector Core).

### Genome targeting and quantification

After electroporation with the RAG2 RNP complex, CD34<sup>+</sup> HSPCs were seeded at a density of 0.4 × 10<sup>6</sup> cells/ml for 24 hrs. Following the incubation period, the cell density was adjusted to 0.25 × 10<sup>6</sup> cells/ml for an additional 24 hrs. In HDR experiments, the cells were transduced with the rAAV6 donor at MOIs of either 12,500 or 25,000 VG/cell within 5 min of electroporation. Cells were cultured at 37 °C, 5% CO<sub>2</sub>, and 5% O<sub>2</sub> in StemSpan™ SFEM II enriched medium as noted above. Flow cytometry analyses were performed on the BD LSRFortessa™ (BD Biosciences), Aria III cell sorter (BD Biosciences), or the Accuri C6 flow cytometer (BD Biosciences). Flow cytometry analysis was performed using the FlowJo Software, Version 10.9 (<https://www.flowjo.com/>).

### IVTD system and immunostaining

CD34<sup>+</sup> HSPCs were cultured in the StemSpan™ T-Cell Generation Kit (STEMCELL Technologies, Inc.). For the first 14 days, cells were cultured in StemSpan™ SFEM II medium containing Lymphoid Progenitor Expansion Supplement in plates pre-coated with the Lymphoid Differentiation Coating Material. Cells were then harvested and re-seeded for an additional 14 days in StemSpan™ SFEM II medium containing the T-Cell Progenitor Maturation Supplement on pre-coated plates. Flow cytometry analysis was conducted on days 14 and 28 of IVTD using the LSR Fortessa™ (BD Biosciences). On day 14 of IVTD, cells were stained with PE/Cy7-anti-CD7 1:20 (clone: CD7-6B7, BioLegend), BV421-anti-CD5 1:20 (clone: UCHT2, BioLegend), PE-anti-CD1a 1:5 (clone: BL6, Beckman Coulter), and APC-anti-NGFR 1:20 (clone: ME20.4, BioLegend) antibodies. On day 28 of IVTD, cells were stained with PE/Cy7-anti-CD4 1:20 (clone: RPA-T4, BioLegend), APC-r700-anti-CD8a 1:20 (clone: RPA-T8, BD Horizon™) BV421-anti-CD3 1:20 (clone: UCHT1, BioLegend), PE-anti-TCR PAN γ/δ 1:5 (clone: IMMU510, Beckman Coulter), and APC-anti-NGFR 1:20 (clone: ME20.4, BioLegend) antibodies. BD Horizon™ Fixable Viability Stain 510 was performed on all collected cells at both time points. Gating strategies were based on fluorescence minus one (FMO) plus isotype control (at equivalent concentration to its antibody pair) samples using the following isotypes: PE/Cy7 Mouse IgG2a κ, (BioLegend), BV421 Mouse IgG1 κ (BioLegend), PE Mouse IgG1 κ (BioLegend), PE/Cy7 Mouse IgG1 κ (BioLegend), APC-r700 Mouse IgG1 κ (BD Biosciences), and APC Mouse IgG1 κ (BioLegend).

### Digital droplet PCR™ (ddPCR™)

Genomic integration quantification for HDR experiments was performed by Digital Droplet PCR™ (ddPCR™, Bio-Rad). DNA was extracted from cell populations using GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific). Each ddPCR reaction contained a HEX reference assay detecting the *CCRL2* gene to quantify the chromosome 3 copy number input<sup>88</sup>. FAM assays for either *KO* (disruption) or *KI* (correction) donors were designed to detect the locus-specific donor integration (assay schematics depicted in Supplementary Figs. 2F, 3E, and 5C). The ddPCR reaction was carried with the following reagents: 10 μl of ddPCR Supermix for Probes No dUTP (Bio-Rad), 1 μl each of FAM and HEX PrimeTime® Standard qPCR Assay (IDT), 1 μl restriction enzyme mix [5 μl EcoRI-HF® (NEB), 2 μl nuclease-free water, 1 μl CutSmart Buffer 10X (NEB)], genomic template DNA, and supplemented to a total of 20 μl with nuclease-free water. Droplet samples were prepared according to the manufacturer's protocol (Bio-Rad) and

40 μl of the droplet output was transferred to a 96-well plate and amplified in a Bio-Rad PCR thermocycler (Bio-Rad) at the following PCR conditions: 1 cycle at 95 °C for 10 min, then 40 cycles of 95 °C for 30 sec and 55 °C for 3 min, followed by 1 cycle at 98 °C for 10 min with a ramp rate of 2.2 °C/sec. Following the PCR, the plate was read in the QX200 Droplet Reader (Bio-Rad) and the data was analyzed using the QuantaSoft, Version 1.7, Regulatory Edition analysis software (Bio-Rad). Primer and probe sequences are listed in Table 1.

### mRNA quantification

RNA was extracted using Direct-zol™ RNA Miniprep Plus (Zymo Research) from differentiated T cells obtained on day 28 of IVTD. cDNA preparation was executed from RNA, using Oligo d(T)23 VN- S1327S (NEB), dNTPs 10 mM (Sigma-Aldrich), and M-MuLV Reverse Transcriptase (NEB). qRT-PCR reactions were conducted using the TaqMan® Fast Advanced Master Mix (Thermo Fisher Scientific) in the StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific). PCR conditions were as follows: uracil-N-glycosylase gene (UNG) incubation at 50 °C for 2 min, polymerase activation at 95 °C for 20 sec, and 40 cycles at 95 °C for 1 sec and at 60 °C for 20 sec. Primer and probe sequences are listed in Table 1.

### TRB and TRG V(D)J assessment

Genomic DNA was extracted using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific) from differentiated T cells on day 28 of IVTD. For TRG assessment via PCR amplification, 12 possible CDR3 clones were amplified using combinations of 4 primers for the V<sub>γ</sub> regions and 3 primers for the J<sub>γ</sub> regions (primer sequences are

**Table 1 | ddPCR and qRT-PCR assay sequences**

RAG2 KO (DNA)	Forward	CTCTCCAGAGCAACAAAGAC
	Reverse	TTTCATGCCTTGC AAAATGG
	Probe	56-FAM/CCTACAGGT/ZEN/ GGGGTCTTTCATTCC/3IABkFQ
RAG2 KI Correction (DNA)	Forward	TCTCCAGAGCAACAAAG
	Reverse	GAAGTTCATGAGGCTGAAG
	Probe	56-FAM/CATAGCCTT/ZEN/AATCCAGCCCG/ 3IABkFQ
RAG1 KO (DNA)	Forward	CAGTGACTTTCAGGATGACCT
	Reverse	CAGCTAGCTTGCCAAACC
	Probe	56-FAM/CCATGCTGG/ZEN/CTGAGGTACC/ 3IABkFQ
CCRL2 (DNA)	Forward	GCTGTATGAATCCAGGTCC
	Reverse	CCTCCTGGCTGAGAAAAAG
	Probe	5HEX/TGTTTCCTC/ZEN/CAGGA- TAAGGCAGCTGT/3IABkFQ
Endogenous RAG2 (RNA)	Forward	CCAAGTGTGACAATTAATACCTG
	Reverse	GACATGTTTATGCTTACATCCAG
	Probe	56-FAM/CATCAGTGA/ZEN/GAAGCCTGGCT- GAATTAAGG/3IABkFQ
dcoRAG2 cDNA (RNA)	Forward	CAAGTGTGACAATTAATACCT
	Reverse	GAAGTTCATGAGGCTGAAG
	Probe	56-FAM/CATAGCCTT/ZEN/AATCCAGCCCG/ 3IABkFQ
Comparison RAG2 (RNA)	Forward	CCAAGTGTGACAATTAATACCTG
	Reverse	GACATAGTTTCTGATGGTACGTAGA
	Probe	56-FAM/TCACGCCTC/ZEN/ TCTGAATCTTTCGGG/3IABkFQ
GAPDH (RNA)	Forward	ACATCGCTCAGACACCATG
	Reverse	TGTAGTTGAGGTCAATGAAGGG
	Probe	56-FAM/AAGGTCGGA/ZEN/GTCAACG- GATTTGGTC/3IABkFQ

**Table 2 | TRG PCR amplification primers for V(D)J recombination assessment**

Jy <sub>1/2</sub>	TACCTGTGACAACCAAGTGTGG
Jy <sub>P</sub>	ACTTACCTGTAATGATAAGCTTT
Jy <sub>P1/2</sub>	TTACCAGGCGAAGTTACTATG
Vy <sub>9.2</sub>	ACCTGGTGAAGTCATACAGTTC
Vy <sub>11</sub>	CTTCCACTTCCACTTTGAAA
Vy <sub>f1</sub>	ACTGGTACCTACACCAGGAGG
Vy <sub>10.2</sub>	AGCATGGGTAAGACAAGCAA

presented in Table 2). The PCR products were run on a 2% agarose gel. For deep sequencing of the TRB and TRG repertoires, the same genomic DNA was amplified using a multiplex master mix from either the LymphoTrack® TRB assay and/or LymphoTrack® TRG assay kits (Invivoscribe, Inc.). The amplicons were purified and sequenced using the MiSeq V2 (500 cycles) kit, with 250 bp paired-end reads (Illumina). The resulting FASTQ files were analyzed by the LymphoTrack Software – MiSeq, Version 2.4.3 (Invivoscribe, Inc.) and by the IMGT®: ImMunoGeneTics Software, Version 1 (The International ImMunoGeneTics Information System®, HighV-QUEST, <http://www.imgt.org>). The analysis of the incidence and clonality of TRB and TRG rearrangement sequences was performed for visual representation by the Treemap Software, Version 2019.9.1 (Macrofocus GmbH). Unique CDR3 sequence and length were determined from the total productive sequences. Lastly, Shannon's H and Simpson's 1-D diversity indices were calculated using the PAST Past4: Paleontological Statistics Software, version 4 by Øyvind Hammer<sup>59</sup>.

### ITR-seq

ITR-seq protocol was adapted from Breton et al.<sup>4</sup>. Amplicons were generated from purified genomic DNA. DNA was sheared to an average size of 500 bp in 130 µl volume using an ME220 focused-ultrasonicator (Covaris, Woburn, MA), purified using AMPure beads (Beckman Coulter, Indianapolis, IN) at a 0.8x ratio, and eluted in 52 µl of IDTE pH 7.5 buffer (Integrated DNA Technologies, Coralville, Iowa, USA). End repair and dA tailing were subsequently performed in a total volume of 60 µl containing 50 µl of eluted DNA, 7 µl of ERAT buffer, and 3 µl of ERAT mix from KAPA HyperPrep Kit (cat#KK8504, Roche, Basel, Switzerland). The mix was incubated at 20 °C for 30 min, 65 °C for 30 min, and then held at 4 °C. Meanwhile, the unique Y-adapters, with molecular index tags were annealed to the common adapter (IDT) (see Supplementary Table 3 for the sequences). Each P5 adapter sequence and the Common Adapter GS were diluted to 100 µM with IDTE pH 7.5 buffer and further diluted to 30 µM with IDTE pH 7.5 buffer. Duplexes were formed in IDTE pH 7.5 for each P5 adapter sequence with the common adapter at an equimolar ratio and a final concentration of 15 µM. The final volume is according to the number of reactions needed from each P5 adapter in the ligation reaction. For annealing, the mixture was heated to 95 °C for 5 min and slowly cooled to room temperature.

After completion of the ERAT reaction, the ERAT products were ligated to the annealed adapter in the following mix: 5 µl of 10 µM annealed A01-A16 Y-adapter, 5 µl of nuclease-free water, 30 µl of Ligation buffer and 10 µl of Ligase (KAPA HyperPrep Kit, cat. no. KK8504, Roche, Basel, Switzerland), and 60 µl of the previous end-repaired and dA-tailed DNA. The ligation program was 20 °C for 15 min, and then hold at 4 °C. DNA was then purified by AMPure beads (Beckman Coulter, Indianapolis, IN) at a 0.7x ratio and eluted in 11 µl of IDTE pH 7.5 buffer. End-repaired Y-adapter-ligated DNA fragments were amplified by PCR using an ITR-specific primer and an adapter-specific primer (A01-A16\_P5\_FWD primer) (see Supplementary Table 3 for the sequences), in the following mix (amounts per sample): 13.4 µl

of nuclease-free water, 3 µl of 10x buffer for Taq Polymerase (MgCl<sub>2</sub>-free, Invitrogen, Carlsbad, CA), 0.6 µl of 10 mM dNTP mix (Bio-Lab LTD, Jerusalem, Israel), 1.2 µl of 50 mM MgCl<sub>2</sub> (Invitrogen, Carlsbad, CA), 0.3 µl of 5 U/µl Platinum Taq polymerase (Invitrogen, Carlsbad, CA), 1 µl of 10 µM GSP\_ITR3. AAV2 primer, 0.5 µl of 10 µM A01-A16\_P5\_FWD primer with the primer number matching the adapter number (e.g., A01\_P5\_FWD primer to be used with A01 Y-adapter), and 10 µl of previously purified DNA. The PCR program was 1 cycle of 95 °C for 5 min 30 cycles of 95 °C for 30 s, 69 °C for 1 min, and 72 °C for 30 s; 1 cycle at 72 °C for 5 min; and 4 °C hold. PCR products were purified using 0.7x AMPure beads (Beckman Coulter, Indianapolis, IN) and eluted in 17 µl of IDTE pH 7.5 buffer.

NGS libraries were prepared by PCR in the following mix (amounts per sample): 7.9 µl of nuclease-free water (Life Technologies, Waltham, MA), 3 µl of 10x buffer for Taq Polymerase (MgCl<sub>2</sub>-free; Invitrogen, Carlsbad, CA), 0.6 µl of 10 mM dNTP mix (Bio-Lab LTD, Jerusalem, Israel), 1.2 µl of 50 mM MgCl<sub>2</sub> (Invitrogen, Carlsbad, CA), 0.3 µl of 5 U/µl Platinum Taq polymerase (Invitrogen, Carlsbad, CA), 0.5 µl of 10 µM A01-A4\_P5\_FWD primer with the primer number matching the adapter number, 1.5 µl of 10 µM p701–8 primers, and 15 µl of previously purified DNA (including the AMPure beads used in the previous PCR purification step). The PCR program was 1 cycle of 95 °C for 5 min; 10 cycles of 95 °C for 30 s, 75 °C for 2 min (–1 °C/cycle), and 72 °C for 30 s; 15 cycles of 95 °C for 30 s, 69 °C for 1 min, and 72 °C for 30 s; 1 cycle at 72 °C for 5 min; and 4 °C hold. PCR products were purified using 0.7x AMPure beads (Beckman Coulter, Indianapolis, IN), and resuspended in 25 µl of IDTE pH 7.5 buffer. Dual-indexed sequencing libraries were sequenced on an Illumina MiSeq V2 kit generating 2 × 150 bp paired-end reads using custom Index 1 and Read 2 primers (see Supplementary Table 3 for the sequences).

Unique molecular identifier (UMI) tagging and consolidating was performed on raw FASTQ files using the GUIDE-seq UMI module with default parameters (<https://github.com/aryeelab/guideseq>). R1 and R2 files were then trimmed using cutadapt with minimum-length of 15 (<https://github.com/marcelm/cutadapt>). Read pairs were then aligned using the bowtie2 package (<https://github.com/BenLangmead/bowtie2>) in a local mode against the AAV6 genome. Reads with MAPQ > 30 were further aligned to the hg38 human genome, again filtered by MAPQ > 30. Finally, genomic coordinates of validated reads (containing both human genomic sequence and fragments of the ITR sequence) were found using bedtools (<https://github.com/arq5x/bedtools2>).

### Cas9-targeted sequencing

Cas9 enrichment was conducted using Cas9 Sequencing Kit (cat. no. SQK-CS9109) according to the manufacturer's protocol. Two sgRNAs targeting the positive strand upstream to the region of interest, and two sgRNAs targeting the negative strand downstream to the region of interest were designed according to the Oxford Nanopore's guidelines (see Supplementary Table 5 for the sgRNA sequences). sgRNAs, Alt-R S.p. Cas9 nuclease V3, and Alt-R S.p. HiFi Cas9 nuclease V3 were purchased from IDT. The activity of RNP complexes was confirmed by digesting purified PCR amplicons produced in the target cells by specific primers and containing the target sites for the respective sgRNA, as described previously<sup>7</sup>. The sequencing data was generated at the Technion Genomics Center (Faculty of Medicine, Technion, Haifa, Israel) on the Oxford Nanopore MinION device (MinKNOW software v.22.12.7), using R9.4.1 flow cell (Oxford Nanopore, cat. no. FLO-MIN106D). Raw Fast5 files were base called by Guppy Basecalling Software (v6.4.8). FASTQ files were aligned to the hg38 human genome using minimap2 with default parameters<sup>6</sup>. Reads that were mapped to the RAG2 genomic locus were extracted and manually classified by three individuals, using Basic Local Alignment Search Tool (BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) as well as SnapGene software (GSL Biotech LLC, San Diego, CA, USA).

## Statistics and reproducibility

All replicates in this work were conducted on biologically unique CD34<sup>+</sup> HSPCs from different donors. There are no technical replicates presented in these data. Statistical analyses of the data were conducted using an unpaired, one-sided Mann-Whitney test for rank comparison. We found that while the results were reproducible, as expected, donor-to-donor variation was detected in the experiments. Thus, we designed our study to have  $N = 3$  or more for any given analysis. No statistical method was used to predetermine the sample size. No data was excluded unless contamination occurred over the course of the 28 days in the IVTD system. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

## Reporting summary

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

## Data availability

The TRB and TRG sequencing data as well as the ITR-seq and ONT long-read sequencing data generated in this study have been deposited in the Sequence Read Archive (SRA) database under accession number: [PRJNA926613](https://www.ncbi.nlm.nih.gov/sra/PRJNA926613). Source data are provided with this paper.

## References

- Barron, M. A. et al. Increased resting energy expenditure is associated with failure to thrive in infants with severe combined immunodeficiency. *J. Pediatr.* **159**, 628–32.e1 (2011).
- Picard, C. et al. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J. Clin. Immunol.* **35**, 696–726 (2015).
- Cirillo, E. et al. Severe combined immunodeficiency—an update. *Ann. N. Y. Acad. Sci.* **1356**, 90–106 (2015).
- Notarangelo, L. D., Kim, M. S., Walter, J. E. & Lee, Y. N. Human RAG mutations: biochemistry and clinical implications. *Nat. Rev. Immunol.* **16**, 234–246 (2016).
- Rechavi, E. et al. First Year of Israeli Newborn Screening for Severe Combined Immunodeficiency—Clinical Achievements and Insights. *Front Immunol.* **8**, 1448 (2017).
- Lev, A. et al. Lessons Learned From Five Years of Newborn Screening for Severe Combined Immunodeficiency in Israel. *J. Allergy Clin. Immunol. Pract.* **10**, 2722–2731.e9 (2022).
- Kumrah, R. et al. Genetics of severe combined immunodeficiency. *Genes Dis.* **7**, 52–61 (2019).
- Holtzman, N. A. Newborn screening for severe combined immunodeficiency progress and challenges. *JAMA* **312**, 701–702 (2014).
- Kwan, A. et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA* **312**, 729–738 (2014).
- Schlissel, M. S. Regulating antigen-receptor gene assembly. *Nat. Rev. Immunol.* **3**, 890–899 (2003).
- Kuo, T. C. & Schlissel, M. S. Mechanisms controlling expression of the RAG locus during lymphocyte development. *Curr. Opin. Immunol.* **21**, 173–178 (2009).
- Miyazaki, K. & Miyazaki, M. The Interplay Between Chromatin Architecture and Lineage-Specific Transcription Factors and the Regulation of Rag Gene Expression. *Front Immunol.* **12**, 659761 (2021).
- Yannoutsos, N. et al. The role of recombination activating gene (RAG) reinduction in thymocyte development in vivo. *J. Exp. Med.* **194**, 471–480 (2001).
- Yu, W. et al. Coordinate regulation of RAG1 and RAG2 by cell type-specific DNA elements 5' of RAG2. *Science* **285**, 1080–1084 (1999).
- Verkoczy, L. et al. A role for nuclear factor kappa B/rel transcription factors in the regulation of the recombinase activator genes. *Immunity* **22**, 519–531 (2005).
- Wilson, A., Held, W. & MacDonald, H. R. Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* **179**, 1355–1360 (1994).
- Hao, B. et al. An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J. Exp. Med.* **212**, 809–824 (2015).
- Naik, A. K., Byrd, A. T., Lucander, A. C. K. & Krangel, M. S. Hierarchical assembly and disassembly of a transcriptionally active RAG locus in CD4 + CD8 + thymocytes. *J. Exp. Med.* **216**, 231–243 (2019).
- Miyazaki, K. et al. The transcription factor E2A activates multiple enhancers that drive Rag expression in developing T and B cells. *Sci. Immunol.* **5**, eabb1455 (2020).
- Rommel, P. C., Oliveira, T. Y., Nussenzweig, M. C. & Robbiani, D. F. Correction: RAG1/2 induces genomic insertions by mobilizing DNA into RAG1/2-independent breaks. *J. Exp. Med.* **214**, 2167 (2017).
- Raghavan, S. C., Kirsch, I. R. & Lieber, M. R. Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. *J. Biol. Chem.* **276**, 29126–29133 (2001).
- Nussenzweig, A. & Nussenzweig, M. C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27–38 (2010).
- Teng, G. & Schatz, D. G. Regulation and Evolution of the RAG Recombinase. *Adv. Immunol.* **128**, 1–39 (2015).
- Wayne, J. et al. TCR selection and allelic exclusion in RAG transgenic mice that exhibit abnormal T cell localization in lymph nodes and lymphatics. *J. Immunol.* **153**, 5491–5502 (1994).
- Gennery, A. R. et al. Transplantation of hematopoietic stem cells and long-term survival for primary immunodeficiencies in Europe: entering a new century, do we do better? *J. Allergy Clin. Immunol.* **126**, 602–10.e1-11 (2010).
- Schuetz, C. et al. SCID patients with ARTEMIS vs RAG deficiencies following HCT: increased risk of late toxicity in ARTEMIS-deficient SCID. *Blood* **123**, 281–289 (2014).
- Pai, S.-Y. et al. Transplantation outcomes for severe combined immunodeficiency, 2000–2009. *N. Engl. J. Med.* **371**, 434–446 (2014).
- Cowan, M. J. & Gennery, A. R. Radiation-sensitive severe combined immunodeficiency: The arguments for and against conditioning before hematopoietic cell transplantation—what to do? *J. Allergy Clin. Immunol.* **136**, 1178–1185 (2015).
- Cossu, F. Genetics of SCID. *Ital. J. Pediatr.* **36**, 76 (2010).
- Booth, C., Gaspar, H. B. & Thrasher, A. J. Treating Immunodeficiency through HSC Gene Therapy. *Trends Mol. Med.* **22**, 317–327 (2016).
- Cicalese, M. P. & Aiuti, A. Clinical applications of gene therapy for primary immunodeficiencies. *Hum. Gene Ther.* **26**, 210–219 (2015).
- Dever, D. P. et al. CRISPR/Cas9  $\beta$ -globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384–389 (2016).
- Naldini, L. Ex vivo gene transfer and correction for cell-based therapies. *Nat. Rev. Genet.* **12**, 301–315 (2011).
- de Ravin, S. S. et al. Lentiviral hematopoietic stem cell gene therapy for X-linked severe combined immunodeficiency. *Sci. Transl. Med.* **8**, 335ra57 (2016).
- Marquez Loza, L. I., Yuen, E. C. & McCray, P. B. Lentiviral Vectors for the Treatment and Prevention of Cystic Fibrosis Lung Disease. *Genes (Basel)* **10**, 218 (2019).
- Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia* **32**, 1529–1541 (2018).
- Kang, H. J. et al. Retroviral gene therapy for X-linked chronic granulomatous disease: results from phase I/II trial. *Mol. Ther.* **19**, 2092–2101 (2011).
- Moratto, D. et al. Long-term outcome and lineage-specific chimerism in 194 patients with Wiskott-Aldrich syndrome treated by

- hematopoietic cell transplantation in the period 1980-2009: an international collaborative study. *Blood* **118**, 1675–1684 (2011).
39. Boztug, K. et al. Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N. Engl. J. Med.* **363**, 1918–1927 (2010).
40. Pavel-Dinu, M. et al. Gene correction for SCID-X1 in long-term hematopoietic stem cells. *Nat. Commun.* **10**, 1634 (2019).
41. Candotti, F. et al. Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans. *Blood* **120**, 3635–3646 (2012).
42. Howe, S. J. et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* **118**, 3143–3150 (2008).
43. Nota Informativa Importante—Strimvelis®.
44. Markt, S. et al. Intrabone hematopoietic stem cell gene therapy for adult and pediatric patients affected by transfusion-dependent  $\beta$ -thalassemia. *Nat. Med.* **25**, 234–241 (2019).
45. Brunetti-Pierri, N. Safety and efficacy of gene-based therapeutics for inherited disorders. (Springer International Publishing, 2017).
46. Garcia-Perez, L. et al. Successful Preclinical Development of Gene Therapy for Recombinase-Activating Gene-1-Deficient SCID. *Mol. Ther. Methods Clin. Dev.* **17**, 666–682 (2020).
47. Bak, R. O., Dever, D. P. & Porteus, M. H. CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat. Protoc.* **13**, 358–376 (2018).
48. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
49. Chen, X. & Gonçalves, M. A. F. V. Engineered Viruses as Genome Editing Devices. *Mol. Ther.* **24**, 447–457 (2016).
50. 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. *CRISPR J.* **5**, 80–94 (2022).
51. Porteus, M. Genome Editing: A New Approach to Human Therapeutics. *Annu Rev. Pharm. Toxicol.* **56**, 163–190 (2016).
52. Powell, S. K., Rivera-Soto, R. & Gray, S. J. Viral expression cassette elements to enhance transgene target specificity and expression in gene therapy. *Discov. Med.* **19**, 49–57 (2015).
53. Cromer, M. K. et al. Gene replacement of  $\alpha$ -globin with  $\beta$ -globin restores hemoglobin balance in  $\beta$ -thalassemia-derived hematopoietic stem and progenitor cells. *Nat. Med.* **27**, 677–687 (2021).
54. Gray, S. J. et al. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Hum. Gene Ther.* **22**, 1143–1153 (2011).
55. Donello, J. E., Loeb, J. E. & Hope, T. J. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. *J. Virol.* **72**, 5085–5092 (1998).
56. Iancu, O. et al. Multiplex HDR for disease and correction modeling of SCID by CRISPR genome editing in human HSPCs. *Mol. Ther. Nucleic Acids* **31**, 105–121 (2023).
57. Monroe, R. J., Chen, F., Ferrini, R., Davidson, L. & Alt, F. W. RAG2 is regulated differentially in B and T cells by elements 5' of the promoter. *Proc. Natl Acad. Sci. USA.* **96**, 12713–12718 (1999).
58. Fong, I. C., Zarrin, A. A., Wu, G. E. & Berinstein, N. L. Functional analysis of the human RAG 2 promoter. *Mol. Immunol.* **37**, 391–402 (2000).
59. Jankovic, M., Casellas, R., Yannoutsos, N., Wardemann, H. & Nussenzweig, M. C. RAGs and regulation of autoantibodies. *Annu Rev. Immunol.* **22**, 485–501 (2004).
60. Yannoutsos, N. et al. A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. *Nat. Immunol.* **5**, 443–450 (2004).
61. Agnieszka, Ł. et al. Ikaros and RAG-2-mediated antisense transcription are responsible for lymphocyte-specific inactivation of NWC promoter. *PLoS One* **9**, e106927 (2014).
62. Kisielow, P., Miazek, A. & Cebrat, M. NWC, a new gene within RAG locus: could it keep GOD under control? *Int J. Immunogenet.* **35**, 395–399 (2008).
63. Amit, I. et al. CRISPECTOR provides accurate estimation of genome editing translocation and off-target activity from comparative NGS data. *Nat. Commun.* **12**, 3042 (2021).
64. Shapiro, J. et al. Increasing CRISPR Efficiency and Measuring Its Specificity in HSPCs Using a Clinically Relevant System. *Mol. Ther. Methods Clin. Dev.* **17**, 1097–1107 (2020).
65. Wang, J. et al. Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat. Biotechnol.* **33**, 1256–1263 (2015).
66. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **31**, 317–334 (2017).
67. Bijlani, S., Pang, K. M., Sivanandam, V., Singh, A. & Chatterjee, S. The Role of Recombinant AAV in Precise Genome Editing. *Front Genome Ed.* **3**, 799722 (2021).
68. Breton, C., Clark, P. M., Wang, L., Greig, J. A. & Wilson, J. M. ITR-Seq, a next-generation sequencing assay, identifies genome-wide DNA editing sites in vivo following adeno-associated viral vector-mediated genome editing. *BMC Genom.* **21**, 239 (2020).
69. Gilpatrick, T. et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
70. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **36**, 767–773 (2004).
71. Nelson, C. E. et al. Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nat. Med.* **25**, 427–432 (2019).
72. Gil-Farina, I. et al. Recombinant AAV Integration Is Not Associated With Hepatic Genotoxicity in Nonhuman Primates and Patients. *Mol. Ther.* **24**, 1100–1105 (2016).
73. Gardner, C. L. et al. Gene Editing Rescues In vitro T Cell Development of RAG2-Deficient Induced Pluripotent Stem Cells in an Artificial Thymic Organoid System. *J. Clin. Immunol.* **41**, 852–862 (2021).
74. Themeli, M. et al. iPSC-Based Modeling of RAG2 Severe Combined Immunodeficiency Reveals Multiple T Cell Developmental Arrests. *Stem Cell Rep.* **14**, 300–311 (2020).
75. Chang, C. W. et al. Modeling Human Severe Combined Immunodeficiency and Correction by CRISPR/Cas9-Enhanced Gene Targeting. *Cell Rep.* **12**, 1668–1677 (2015).
76. Brauer, P. M. et al. Modeling altered T-cell development with induced pluripotent stem cells from patients with RAG1-dependent immune deficiencies. *Blood* **128**, 783–793 (2016).
77. Shahzad, M. et al. Outcomes with CD34-Selected Stem Cell Boost for Poor Graft Function after Allogeneic Hematopoietic Stem Cell Transplantation: A Systematic Review and Meta-Analysis. *Transpl. Cell Ther.* **27**, 877.e1–877.e8 (2021).
78. Ferrari, S. et al. Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking. *Nat. Biotechnol.* **38**, 1298–1308 (2020).
79. Wang, C. Y. et al. Analysis of the CDR3 length repertoire and the diversity of T cell receptor  $\alpha$  and  $\beta$  chains in swine CD4+ and CD8+ T lymphocytes. *Mol. Med Rep.* **16**, 75–86 (2017).
80. Hubbard, N. et al. Targeted gene editing restores regulated CD40L function in X-linked hyper-IgM syndrome. *Blood* **127**, 2513–2522 (2016).
81. Pavel-Dinu, M. et al. Genetically Corrected RAG2-SCID Human Hematopoietic Stem Cells Restore V(D)J-Recombinase and Rescue Lymphoid Deficiency. *bioRxiv*. 2022.07.12.499831. (2022).

82. Ylä-Herttuala, S. Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. *Mol. Ther.* **20**, 1831–1832 (2012).
83. Shams, F. et al. Advance trends in targeting homology-directed repair for accurate gene editing: An inclusive review of small molecules and modified CRISPR-Cas9 systems. *Bioimpacts* **12**, 371–391 (2022).
84. Liu, M. et al. Methodologies for Improving HDR Efficiency. *Front Genet.* **9**, 691 (2019).
85. 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).
86. Gray, D. H. et al. Optimizing Integration and Expression of Transgenic Bruton's Tyrosine Kinase for CRISPR-Cas9-Mediated Gene Editing of X-Linked Agammaglobulinemia. *CRISPR J.* **4**, 191–206 (2021).
87. Shapiro, J., Tovin, A., Iancu, O., Allen, D. & Hendel, A. Chemical Modification of Guide RNAs for Improved CRISPR Activity in CD34+ Human Hematopoietic Stem and Progenitor Cells. *Methods Mol. Biol.* **2162**, 37–48 (2021).
88. Gomez-Ospina, N. et al. Human genome-edited hematopoietic stem cells phenotypically correct Mucopolysaccharidosis type I. *Nat. Commun.* **10**, 4045 (2019).
89. Hammer, D. A. T., Ryan, P. D., Hammer, Ø. & Harper, D. A. T. Past: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **4**, 178 (2001).

## Acknowledgements

We would like to thank the members of the Somech and the Hendel Labs for reading the manuscript and providing practical advice. Additionally, we would like to thank D. Russell for providing the pDGM6 plasmid. We give special thanks to the Technion Genomics Center team, and especially Dr. Nitsan Fourier, for excellent technical assistance and support with the long-read sequencing. This study was supported in part by research funding from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (Grant No. 755758 A.H.) as well as the Israel Science Foundation (ISF)—Israel Precision Medicine Partnership (IPMP) (Grant No. 3115/19 A.H.) and Israel Science Foundation (ISF)—Individual Research Grants (Grant No. 2031/19 A.H.).

## Author contributions

D.A., O.K., B.I., and M.R. designed and conducted the experiments; D.A., N.K., and M.R. evaluated, and analyzed the data and performed the

bioinformatics analyses, with the help and guidance of O.I. and Y.N.L.; K.B., and A.N. provided cord blood samples; K.B., Y.N.L., A.N., and R.S. critically reviewed the experiments and provided important advice; A.H. supervised and conceived the research; D.A. and A.H. wrote the manuscript, with contributions and input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-42036-5>.

**Correspondence** and requests for materials should be addressed to Ayal Hendel.

**Peer review information** *Nature Communications* thanks Harald Mikkers, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023