Multiplexed Genetic Engineering of Human Hematopoietic Stem and Progenitor Cells using CRISPR/Cas9 and AAV6

Rasmus O. Bak[1, 3], Daniel P. Dever[1, 3], Andreas Reinisch[2, 3], David Cruz[2], Ravindra Majeti[2, 4] & Matthew H. Porteus[1, 4]

[1]Department of Pediatrics, Stanford University, Stanford, CA 94305, USA.

[2]Department of Medicine, Division of Hematology, Cancer Institute, and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA 94305, USA.

[3]Co-first author

[4]To whom correspondence may be addressed.

Corresponding authors:
Ravi Majeti (rmajeti@stanford.edu) and Matthew H. Porteus (mporteus@stanford.edu)

ABSTRACT

Precise and efficient manipulation of genes is crucial for understanding the molecular mechanisms that govern human hematopoiesis and for developing novel therapies for diseases of the blood and immune system. Current methods do not enable precise engineering of complex genotypes that can be easily tracked in a mixed population of cells. We describe a method to multiplex homologous recombination (HR) in human hematopoietic stem and progenitor cells and primary human T cells by combining rAAV6 donor delivery and the CRISPR/Cas9 system delivered as ribonucleoproteins (RNPs). In addition, the use of reporter genes allows FACS-purification and tracking of cells that have had multiple alleles or loci modified by HR. We believe this method will enable broad applications not only to the study of human hematopoietic gene function and networks, but also to perform sophisticated synthetic biology to develop innovative engineered stem cell-based therapeutics.

INTRODUCTION

The current gold standard method for studying human hematopoietic stem and progenitor cell (HSPC) gene function has been either overexpression or RNAi-mediated knockdown of genes using lentiviral vectors (1, 2). While these methods have provided great insights into HSPC biology, they come with several confounders, such as random integration of the vector into the host genome, unregulated transgene expression, and incomplete gene knockdown (3, 4). More recently, programmable nucleases such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR/Cas9 have been utilized to disrupt genes by the introduction of site-specific DNA double strand breaks (DSBs) that are corrected through non-homologous end-joining (NHEJ) (5-11). This error-prone system creates a heterogeneous mixture of cells with various genotypes of SNPs and small insertions or deletions (INDELs); moreover, not all of the genetic changes from INDELs cause functional gene

disruption as they may preserve the open reading frame and may not change amino acids essential for protein functions (12, 13). In a prior study, defined gene deletions were created in HSPCs using a dual sgRNA approach, however, more than half of the alleles were not modified leading to residual gene expression (8). Another limitation of this prior study is that successfully modified cells were not distinguishable from unmodified wild type (WT) cells, and therefore could not be tracked or isolated as an enriched population. Although the versatility of the CRISPR/Cas9 system allows for simultaneous manipulation at multiple genetic loci in a single cell, multiplexing of NHEJ-based gene editing has mainly been performed in immortalized human cancer cell lines and mouse cells (13-17). Finally, these interesting multiplexed proof-of-concept studies, only used NHEJ-mediated editing and did not harness the power of homologous recombination (HR) to create more sophisticated alterations to the genome at multiple alleles and/or loci.

Here, we report an HR-mediated genome engineering method in human HSPCs and T cells that overcomes these limitations and enables the generation and enrichment of HSPC or T cell populations with complete gene knockout or gene replacement at multiple genetic loci. This method has the power to reveal functional gene networks during hematopoiesis and immune system disease pathogenesis and could be combined with the concepts of synthetic biology to create novel stem cell based therapeutics.

RESULTS

Enriching HSPCs with targeted integration

We and others have previously shown that HR in human HSPCs can be efficiently induced by site-specific nucleases in combination with homologous donor DNA delivered as single-stranded oligonucleotides (ssODNs), integration-defective lentiviral vectors (ÍDLVs), or by recombinant adeno-associated virus serotype 6 (rAAV6) vectors (18-22). We previously showed targeted integration in the beta-globin gene (*HBB*) by combining delivery of Cas9 protein pre-

complexed with chemically modified sgRNAs (RNP) and delivery of an AAV6 donor. After successful on-target integration of a reporter transgene, FACS-based sorting of transgene reporter^{high}-expressing HSPCs was used to purify an HSPC population with >90% targeted integration that displayed long-term repopulation capacity in NSG mice (18). To extend this method beyond the *HBB* locus for therapeutic genome editing approaches of hemoglobinopathies, we tested six additional loci for their potential to be modified through HR by CRISPR/Cas9 in combination with AAV6-derived donor delivery. These genes are associated with hematopoiesis, hematopoietic malignancies, or safe harbor sites and include: interleukin-2 receptor gamma chain (*IL2RG*), chemokine (C-C motif) receptor 5 (*CCR5*), runt-related transcription factor 1 isoform c (*RUNX1c*), additional sex combs like 1 (*ASXL1*), stromal antigen 2 (*STAG2*), and adeno-associated virus integration site 1 (*AAVS1*) (23-27). Following electroporation with Cas9 RNP, containing a chemically-modified sgRNA targeting a single site in the selected locus, and transduction with an rAAV6 donor vector carrying homology arms for the targeted site and an expression cassette encoding a fluorescent reporter gene (**Figure 1–figure supplement 1a**), we observed at early time points (day 4) a cell population with increased fluorescence intensity detectable by flow cytometry (reporter^{high} cells) compared to cells receiving only the rAAV6 donor without electroporation of Cas9 RNP (reporter^{low}) (**Fig 1a and Supplementary file 1a)**. For cells targeted at either *CCR5* or *IL2RG,* reporter^{high}, reporter^{low}, and reporter^{neg} populations were sorted at day 4 post-electroporation and cultured up to 22 days. Reporter^{high} populations remained 99.2 ± 0.7% reporter positive (**Fig 1b**) while sorted reporter^{low} and reporter^{neg} populations were 29.3 ± 5.4% and 0.6 ± 0.2% reporter positive, respectively. Dividing the reporter^{low} cells into three sub fractions based on fluorescence intensity revealed that GFP intensity at day 4 post-electroporation positively correlated with the propensity for maintaining GFP expression at day 20 (**Figure 1–figure supplement 1b-c**). In addition, single reporter^{high} cells were plated in methylcellulose to assess integration events at the clonal level. Targeted HSPCs formed a mix of myeloid (CFU-M/GM) and erythroid colonies (BFU-E, CFU-E)

indicating that they retained HSPC function. 'In-Out PCR' (one donor-specific primer and one locus-specific primer outside homology arms) on genomic DNA (gDNA) from single cell-derived methylcellulose colonies confirmed that 99%, 92%, and 100% of reporter^{high} HSPCs targeted at *CCR5* (338 clones analyzed), *IL2RG* (117 clones analyzed), and *RUNX1* (36 clones analyzed), respectively, had at least a monoallelic targeted integration (**Fig 1c and Figure 1–figure supplement 2)**. Analyses of clones with only mono-allelic integration showed gene-specific differences in the modification of the non-integrated alleles ranging from 38% INDELs for *IL2RG* to 89% INDELs for *CCR5* to 88% INDELs for *RUNX1*, among which the majority was gene-disrupting (**Figure 1–figure supplement 2 and Supplementary file 1b**). Collectively, these data indicate that the observed log-fold transgene expression shift following rAAV6 and RNP delivery is due to HR at the intended locus and that reporter expression can be used to enrich gene-targeted HSPCs.

To evaluate the applicability of this technology in a biologically relevant setting we decided to modify the cohesin complex member, *STAG2,* in primary CD34^{+} HSPCs. The cohesin complex has previously been shown to play an essential part in maintaining normal erythroid differentiation potential of hematopoietic stem and progenitor cells (26, 28, 29). Since the *STAG2* gene is located on the human X chromosome, single-allele integration of a fluorescent reporter in male cells would be sufficient to fully knock out the gene. As expected, Cas9 RNP combined with rAAV6 donor transduction resulted in the generation of a reporter^{high} population that could be sorted for subsequent differentiation experiments. Single cell methylcellulose assays of reporter^{high} cells revealed an almost complete loss in the capacity to form erythroid colonies compared to cells that had only been exposed to rAAV6 and not Cas9 RNP, and also compared to cells with targeted integration at the *AAVS1* locus (**Figure 1d**). These proof-of-concept studies provide evidence that gene-specific enrichment of reporter^{high} cells can be used to study HSPC gene function.

Biallelic targeted integration in HSPCs

To determine if this method could be used to enrich HSPCs with biallelic gene disruption, necessary for complete functional gene knockout, we targeted the *ASXL1* gene and simultaneously provided GFP and BFP-encoding rAAV6 donors. Four days after electroporation and transduction, 10.4% of cells were double positive for GFP^{high} and BFP^{high} compared to 0.2% for the AAV only sample (**Fig 2a**). Similarly, double-positive populations were apparent when targeting three other genes (*RUNX1*, *HBB*, and *CCR5)* with two rAAV6 donors with various color combinations (**Figure 2–figure supplement 1 and Supplementary file 1c**). Double-positive cells sorted at day 4 after electroporation remained 94% double-positive for more than two weeks in culture (**Fig. 2b**). 'In-out PCR' on gDNA from single cell-derived methylcellulose clones confirmed on-target integration of one transgene into one allele and the other transgene into the second allele (**Fig. 2c**). We next tested if the biallelic targeting approach could be extended to another blood cell type and therefore targeted primary human T cells for biallelic HR at *CCR5.* After electroporation with *CCR5*-targeting Cas9 RNP followed by transduction with GFP and mCherry *CCR5* rAAV6 donors, a GFP^{high}/$mCherry^{high}$ double-positive population was observed, indicative of biallelic integration at the *CCR5* gene (**Fig. 2d**). No significant toxicity was associated with biallelic targeting in T cells (**Figure 2–figure supplement 2**). Overall, these results demonstrate the utility of using rAAV6, Cas9 RNP, and FACS to enrich for primary human HSPCs and T cells that have undergone biallelic homologous recombination, which may have applications for studying hematological and immunological diseases or generating HSPC or T cell therapeutics that require gene modifications or gene knockout at both alleles.

Simultaneous HR-mediated Targeting of Two Genes (Di-Genic) in HSPCs

The vast majority of hematopoietic functions and immune diseases are governed by complex, polygenic networks (30). To potentially study gene-gene interactions and/or generate cell therapeutics with HR modifications at two separate genes, we tested whether our

methodology could facilitate simultaneous di-genic (two different genes) HR in HSPCs. We therefore co-delivered *HBB*-tdTomato and *IL2RG*-GFP rAAV6 donors with Cas9 RNP targeting both genes. This strategy produced 10.2% double positive GFP^{high}/tdTomato^{high} HSPCs compared to 0.1% for the AAV only control sample (**Fig. 3a**). We also generated double reporter^{high} positive populations when testing other combinations of di-genic HR (*IL2RG*/*CCR5,* *RUNX1*/*ASXL1*, and *HBB*/*CCR5*) (**Figure 3–figure supplement 1 and Supplementary file 1c**). Again, double reporter^{high} positive cells sorted at day 4 post-electroporation remained 94% double positive for 15 days in culture (**Fig. 3b**). 'In-Out PCR' on double positive methylcellulose myeloid and erythroid clones showed on-target integration at both loci in 88% of clones (57 clones analyzed) (**Fig. 3c and 3d**).

Since the combination of two sgRNAs has previously been used to create and study oncogenic translocations (31), and multiplexed TALEN-mediated gene editing in primary human T cells led to translocation frequencies between the two targeted genes of 0.01% - 1% with monocentric translocations occurring most frequently (32), we assessed if our di-genic targeting scheme would enrich for translocations after purification of dual-reporter positive cells. Therefore, we analyzed one of the monocentric translocations between *HBB* and *AAVS1* (**Figure 3-figure supplement 2a**). We targeted *HBB* and *AAVS1* with a GFP and BFP reporter, respectively, and sorted the four different populations (double negative, single positives (each gene), and double positive) seven days after targeting (**Figure 3e, left panel**). INDEL rates at *HBB* and *AAVS1* were comparable among all four sorted populations, with a small enrichment of INDELs in the three populations positive for the reporter (**Figure 3e, middle panel**). Droplet digital PCR (ddPCR) quantification of the translocation showed frequencies ranging from 0.14 - 0.28%, and importantly, no evidence of enrichment of the translocation was observed in the population sorted for di-genic targeting (**Figure 3e, right panel and Figure 3-figure supplement 2c**). Cloning and sequencing of PCR products spanning the translocation showed a wide variety of

translocation junctions derived from different DNA end-processing products (**Figure 3-figure supplement 2b**).

To confirm that HSPCs with long-term and multi-lineage engraftment potential were targeted, we again targeted *HBB* and *AAVS1* with a GFP and BFP reporter, respectively, and transplanted the four different sorted populations into immune-compromised NSG mice (**Fig. 3f**). 12 weeks after transplantation, human multi-lineage engraftment was evident in the bone marrow of the transplanted mice of all four groups (**Fig. 3g and Figure 3-figure supplement 3**).

Collectively, these data show that human HSPCs that have undergone di-genic HR are not enriched for translocations, and maintain their multi-lineage colony forming capacity and long-term engraftment potential.

Multiplexed homologous recombination in HSPCs

We next tested if we could combine the di-genic and biallelic targeting approach to simultaneously target both alleles of *ASXL1* (GFP and mCherry) as well as both alleles of *RUNX1c* (BFP and E2-Crimson) (tetra-allelic) (for schematic see **Figure 4-figure supplement 1a**). Delivery of Cas9 RNPs targeting both genes followed by transduction of four rAAV6 donors gave rise to 1.1% GFP^{high}/$mCherry^{high}$/BFP^{high}/$E2Crimson^{high}$ quadruple-positive cells (**Fig 4a and Figure 4-figure supplement 1b-c**). A similar quadruple-positive population was evident when targeting all four combined alleles of *HBB* and *RUNX1c* (**Figure 4-figure supplement 1e-h and Supplementary file 1e**). Mixed, myeloid, and erythroid colonies were formed at frequency and ratio comparable to AAV only controls (**Fig. 4b**). Genotyping of colonies revealed on-target integration at both alleles at both loci in 78% of clones (73 clones analyzed) (**Fig. 4c**). Flow-cytometric analysis of individual colonies confirmed expression of all four reporters (BFP/GFP/mCherry/E2Crimson) at high levels (**Figure 4-figure supplement 1d)**. The total number of genetic changes in this enriched population, which could be used for synthetic biology purposes is six: two endogenous genes inactivated (both alleles of each gene) plus the addition

of four different transgenes (represented in our experiment by four genes encoding different fluorescent proteins). Thus, this methodology could be used for studying interaction of genes that need both copies disrupted to lose function, such as tumor suppressor genes.

Multi-genic HR in HSPCs would allow for the characterization of functional gene networks during human hematopoiesis (33). To validate that our methodology could multiplex HR in HSPCs in more than two genes simultaneously, we electroporated HSPCs with RNPs targeting *HBB*, *CCR5*, and *IL2RG*, and then transduced them with gene-specific rAAV6s (*HBB*-tdTomato, *CCR5*-tNGFR, *IL2RG*-GFP) (for schematic see **Figure 4-figure supplement 2a**). At day 4 post-electroporation, 4.1% of HSPCs were triple-positive (**Fig. 4d and Figure 4-figure supplement 2b**). 'In-Out PCR' on gDNA from myeloid and erythroid colonies derived from this population showed that 78% (27 clones analyzed) had an integration event at all 3 loci, indicating at least mono-allelic integrations at each targeted locus (**Fig. 4e**). Further analyses showed that 85% of these clones with tri-genic integrations were modified on all alleles either by biallelic integration or INDELs on the non-integrated allele that were mostly disruptive (**Supplementary file 1d**). These data confirm that the methodology can efficiently enrich for HSPCs with multiplexed HR. Targeting at another combination of three genes *(RUNX1/HBB/ASXL1*) showed 2.9% triple-positive cells (**Figure 4-figure supplement 2c-e),** and collectively, tri-genic targeting experiments yielded an average of 4.5% triple-positive cells, with the highest frequency of 14% (*N* = 5) (**Supplementary file 1e**). To test if multiplexing HR caused cellular senescence or more cell death than mono or di-genic targeting in HSPCs, we evaluated cell death and apoptosis rates at day 3 post-targeting and proliferation for up to 10 days post-targeting (corresponding to 7 days post-sorting). We observed similar proliferation rates comparing modified and unmodified cells (data not shown) and only a minor, non-statistically significant decrease in cell viability (p=0.333) when targeting three genes compared to one (**Figure 4-figure supplement 3**). Finally, we targeted HSPCs for tetra-genic HR (*HBB*, *CCR5*, *ASXL1*, *RUNX1*) and found after four days in culture that 1% of cells were reporter^{high} positive for all four reporters (**Fig. 4f**). Strikingly, 41% -

71% of HSPCs with tri-genic HR had undergone tetra-genic HR, suggesting that HR events at different genes may not be independent of each other, in contrast to recent findings for multiplexed NHEJ (13). Targeting the same four genes with other combinations of reporter genes gave 0.41% and 0.78% tetra-genic targeting frequencies in the total cell population (**Supplementary file 1e**). Because rAAVs can be captured at DSBs via NHEJ (34), we performed experiments that aimed to detect the frequency of capture events by including a non-homologous rAAV donor in targeting experiments. We found that 89% - 98% of reporter^{high} cells were derived from on-target homologous recombination, confirming a relatively low rate of AAV capture (**Figure 4-figure supplement 4**).

DISCUSSION

Table 1 summarizes the HR multiplex experiments (seven total genes targeted) and shows that by using Cas9 RNP, rAAV6, and flow cytometry-based sorting, we can reproducibly generate HSPC populations that have undergone HR events at multiple loci. For synthetic biology purposes, the tetra-genic targeting method, for example, can generate an enriched population of cells with 8 genetic modifications: the knockout of at least a single allele of four different genes while introducing four different transgenes (in this proof-of-concept we used three fluorescent protein reporter genes and one biologically inert cell surface marker (tNGFR) that has been previously used in human clinical trials to track genetically modified hematopoietic stem cells over the course of decades). Our approach to studying gene function in human HSPCs has several advantages over lentiviral-based approaches because it enables: **(1)** multigenic targeted integration (at least four genes), **(2)** enrichment of highly pure edited populations, **(3)** the ability to trace cells with a specific genotype, **(4)** enrichment of a population with biallelic targeting of at least two genes, and **(5)** fluorescent protein-based hematopoietic cell lineage tracing. Our methodology has the potential to advance the biological understanding of gene functions in canonical HSC processes, including self-renewal, differentiation, and engraftment, all of which

248 are critical aspects of fundamental stem cell biology and may augment the efficacy of stem cell
249 based therapeutics.

250 By knocking in four different transgenes into four different genes, the method generates
251 four gene disruptions and four gene additions. However, the use of multiple sgRNAs also
252 increases the chances for off-target effects and chromosomal translocations. By looking for
253 monocentric translocations between two genes (*HBB and AAVS1*), we observed low levels of
254 translocation events similar to previously published studies (32). Such effects are likely sgRNA
255 and target gene-specific and need to be assessed on a case-by-case basis. The observed tetra-
256 genic targeting efficiencies at >0.5% are high enough to be experimentally useful, and though
257 some applications may be restricted by HSPC source and starting cell numbers, our targeting
258 methodology may be combined with recent advances in HSPC expansion protocols (35-38) or
259 with transplantation into a humanized bone marrow ossicle xenotransplantation model, which
260 supports higher engraftment levels compared to a standard NSG model (39). By using reporters
261 as transgenes, one can both enrich and track the modified cells, and by using a transgene
262 cassette in which a potentially biologically active transgene is linked through a 2A peptide or
263 IRES to a reporter gene, one can enrich and track cells that could have up to four different new
264 potentially bioactive genes expressed. Additionally, we and others have recently demonstrated
265 the feasibility of knocking in a cDNA immediately after the start codon of the gene, thereby
266 maintaining endogenous regulatory control over gene expression (18, 40, 41). This provides a
267 genetic engineering toolbox where different types of alleles (WT, knockout, mutant cDNA forms)
268 are fluorescently tagged and can be enriched or tracked in a population with mixed allele
269 combinations. One potential caveat is the requirement for reporter gene expression and the fact
270 that cells must be cultured for 2-3 days until reporter gene expression is detectable and cells can
271 be sorted. Even though we have not detected any obvious negative impact in this or previous
272 studies (18, 42), future studies may further investigate and optimize *ex vivo* culturing conditions,

as well as promoter and reporter choice for minimal impact on biology and repopulation potential of edited HSPCs.

Our methodology could be used for the characterization of gene interactions during blood and immune system disease pathogenesis. For example, functional knockouts can be created at one gene (e.g. reporter knock-in into tumor suppressor gene), while introducing disease-causing polymorphisms at another gene (cDNA expression cassette knock-in into proto-oncogene) (see **Figure 4-figure supplement 5** for schematic). For example, Zhao *et al*., showed that the loss of p53 cooperates with the *Kras*^{G12D} mutation to promote acute myeloid leukemia (AML) in mouse HSPCs using retroviral methodology (43). Our system could be used to address whether these findings can be translated to human HSPCs by achieving site specific HR that would simultaneously knock out a tumor suppressor (e.g. *TP53*) and drive mutant *KRAS* under endogenous regulatory conditions, instead of using strong constitutive exogenous viral promoters with little control over proviral copy number and heterogeneity of transgene expression. However, in cDNA knock-in experiments, proper expression should always be validated since elements in the adjacent reporter expression cassette or the lack of UTRs and introns could influence cDNA expression (44). We also show biallelic integration in primary human T cells at *CCR5*, which could be therapeutically applicable for engineering HIV-resistance, where biallelic knockout of *CCR5* could be combined with expression of different HIV restriction factors (45). Additionally, this approach could be useful to extend recently published studies showing high potency of chimeric antigen receptors (CARs) that were site-specifically integrated into the TRAC gene using CRISPR and AAV6 in primary human T cells (46). Multiplexed gene editing may be used to knock-in different CARs or co-stimulatory ligands into genes that are desirable to knock-out in CAR T cell therapy. We anticipate in the future that multiplexed HR mediated cell engineering will facilitate even more sophisticated uses of synthetic biology based stem cell therapeutics than the examples we have given. Our methodology should

also be widely applicable to other cell types of the hematopoietic system besides HSPCs and T cells, and even to cells of non-hematopoietic origin.

In conclusion, we anticipate that this method will be applicable to studying human hematopoiesis and immune system disease pathogenesis through multiplexed, site-specific genome engineering by HR, which has the potential to lead to new discoveries in human hematopoietic stem cell biology.

MATERIALS AND METHODS

AAV vector production

AAV vector plasmids were cloned in the pAAV-MCS plasmid (Agilent Technologies, Santa Clara, CA) containing ITRs from AAV serotype 2 (AAV2). *CCR5*, *IL2RG*, *HBB*, *RUNX1*, *ASXL1*, and *CXCL12* vectors contained an SFFV promoter, a reporter gene such as tNGFR, MaxGFP (or Citrine), BFP, mCherry, tdTomato or E2Crimson and BGH polyA. MaxGFP and Citrine are referred to as GFP throughout. For translocation and NSG transplantation experiments, a UbC promoter (approx. 1200bp) was used in the *HBB* donor instead of an SFFV promoter. For the T cell experiments, donors carried an EF1α promoter (approx. 1200bp). The homology arms for *IL2RG, ASXL1,* and *CCR5* were 800bp, whereas left and right homology arms for *HBB* were 540bp and 420bp, respectively. The homology arms for *RUNX1, STAG2,* and AAVS1 were 400bp. CCR5 donors used in T cell experiments expressed Citrine or mCherry from the PGK promoter and contained 400bp homology arms. rAAV6 vectors were produced as described with a few modifications (47). Briefly, 293FT cells (Life Technologies, Carlsbad, CA, USA) were seeded at $13x10^6$ cells per dish in ten 15-cm dishes one day before transfection. Each 15-cm dish was transfected using standard PEI transfection with 6 μg ITR-containing plasmid and 22 μg pDGM6 (gift from David Russell, University of Washington, Seattle, WA, USA), which contains the AAV6 cap genes, AAV2 rep genes, and adenovirus 5 helper genes. Cells were incubated for 72 hrs until rAAV6 was harvested from cells by three freeze-thaw cycles followed by a 45 min incubation with TurboNuclease (Abnova, Heidelberg, Germany) or Benzonase (Thermo Fisher) at 250 U/mL. AAV vectors were purified on an iodixanol density gradient by ultracentrifugation at 48,000 rpm for 2.25 hrs at 18 °C. AAV vectors were extracted at the 58-40% iodixanol interface and dialyzed three times in PBS with 5% sorbitol in the last dialysis using a 10K MWCO Slide-A-Lyzer G2 Dialysis Cassette (Thermo Fisher Scientific, Santa Clara, CA, USA). Vectors were added pluronic acid to a final concentration of 0.001%, aliquoted, and then stored at -80°C until

further use. rAAV6 vectors were titered using quantitative PCR to measure number of vector genomes as described before (48).

CD34+ hematopoietic stem and progenitor cells

Frozen $CD34^+$ HSPCs derived from mobilized peripheral blood or cord blood were purchased from AllCells (Alameda, CA, USA) and thawed according to manufacturer's instructions. Fresh $CD34^+$ HSPCs from cord blood were acquired from donors under informed consent via the Binns Program for Cord Blood Research at Stanford University and used without freezing. Fresh $CD34^+$ HSPCs from bone marrow were obtained from Stanford BMT Cell-Therapy Facility after informed consent. $CD34^+$ cells were isolated using a human CD34 MicroBead Kit (Miltenyi Biotec, San Diego, CA, USA). Generally, CB-derived HSPCs perform better in HR experiments. $CD34^+$ HSPCs were cultured in stem cell retention media consisting of StemSpan SFEM II (Stemcell Technologies, Vancouver, Canada) supplemented with SCF (100 ng/ml), TPO (100 ng/ml), Flt3-Ligand (100 ng/ml), IL-6 (100 ng/ml), UM171 (Stemcell Technologies) (35nM) and StemRegenin1 (0.75 mM). Mycoplasma contamination testing was not performed. Cells were cultured at 37°C, 5% CO_2, and 5% O_2.

T cell isolation and culturing

Primary human $CD3^+$ T cells were isolated from buffy coats obtained from the Stanford School of Medicine Blood Center using a human T Cell Isolation Kit (Miltenyi) according to manufacturer's instructions. Cells were cultured in X-VIVO 15 (Lonza, Walkersville, MD, USA) containing 5% human serum (Sigma-Aldrich, St. Louis, MO, USA), 100 IU/ml human rIL-2 (Peprotech, Rocky Hill, NJ, USA) and 10 ng/ml human rIL-7 (BD Biosciences, San Jose, CA, USA). T cells were activated directly after isolation with immobilized anti-CD3 antibody (clone: OKT3, Tonbo Biosciences, San Diego, CA, USA) and soluble anti-CD28 antibody (clone: CD28.2, Tonbo

Biosciences) for 72h. Mycoplasma contamination testing was not performed. T cells were cultured at 37 °C, 5% CO_2, and ambient oxygen levels.

Electroporation and transduction of cells

All synthetic sgRNAs were purchased from TriLink BioTechnologies (San Diego, CA, USA). sgRNAs were chemically modified with three terminal nucleotides at both the 5′ and 3′ ends containing 2′ O-Methyl 3′ phosphorothioate and HPLC-purified. The genomic sgRNA target sequences with PAM in bold) were: *HBB*: 5'-CTTGCCCCACAGGGCAGTAA**CGG**-3', *CCR5*: 5'-GCAGCATAGTGAGCCCAGAA**GGG**-3', *IL2RG*: 5'-TGGTAATGATGGCTTCAACA**TGG**-3', *RUNX1c*: 5'-TACCCACAGTGCTTCATGAG**AGG**-3' *ASXL1*: 5'-ACAGATTCTGCAGGTCATAG**AGG**-3', *STAG2:* 5'-AGTCCCACATGCTATCCACA**AGG**-3', AAVS1: 5'-GGGGCCACTAGGGACAGGAT**TGG**-3'. Cas9 protein was purchased from Life Technologies and Integrated DNA Technologies. Cas9 RNP was made by incubating protein with sgRNA at a molar ratio of 1:2.5 at 25˚C for 10 min immediately prior to electroporation into $CD34^+$ HSPCs or T cells. $CD34^+$ HSPCs were electroporated 1 – 2 days after thawing or isolation. T cells were electroporated three days following activation. Both $CD34^+$ HSPCs and T cells were electroporated using the Lonza Nucleofector 2b (program U-014) or 4D (program EO-100) (we have not detected any device-specific differences in electroporation efficiencies) and the Human T Cell Nucleofection Kit (VPA-1002, Lonza) with the following conditions: $5x10^6$ cells/ml, 150-300 µg/ml Cas9 protein complexed with sgRNA at 1:2.5 molar ratio. Following electroporation, cells were incubated for 15 min at 37˚C after which they were added rAAV6 donor vectors (generally at an MOI (vector genomes/cell) of 50,000-100,000 for each gene). A mock-electroporated control was included in most experiments where cells were handled the same and was electroporated in the same electroporation buffer, but without Cas9 RNP. For experiments targeting multiple loci, electroporation volume and cell numbers were kept the same as stated above, and 150-300 µg/ml Cas9 RNP and MOIs of 50,000-100,000 were used for each

targeted locus, but with no more than a total of 60ug Cas9 per electroporation and 200,000 vector genomes/cell. All AAV vectors were added simultaneously and directly to the cell culture after which the cells were transferred to the incubator without further manipulation. AAV volume was kept less than 20% of the total culturing volume and medium was either supplemented or replaced with fresh medium after overnight culture.

Measuring multiplexed targeted integration of fluorescent and tNGFR donors

Reporter^{high} expression was measured by flow cytometric analyses after 3-4 days post-electroporation and transduction using gates for multiplexed targeted integration set so that 'AAV only' samples (no nuclease) were less than 1% since previous data (not presented) have shown that after ~14 days in culture the frequency of reporter^{+} cells (from persistent episomal expression, random integration, and/or non-nuclease mediated HR) is generally less than 1%. The truncated NGFR receptor (tNGFR) where the cytoplasmic intracellular signaling domain is removed and is signaling incompetent, solely served the purpose of a reporter for targeted CD34^{+} HSPCs in indicated experiments (49). Targeted integration of a tNGFR expression cassette was measured by flow cytometry of cells stained with APC-conjugated anti-human CD271 (NGFR) antibody (clone: ME20.4, BioLegend, San Diego, CA). For enriching of reporter^{high} populations, cells were sorted on a FACS Aria II SORP using DAPI, PI (both Thermo Fisher, 1µg/ml) or LIVE/DEAD Fixable Cell Stain Kit (Life Technologies) to discriminate live and dead cells according to manufacturer's instructions.

Scoring, FACS-analysis, and Genotyping of Methylcellulose colonies

Single reporter^{high} cells were either single-cell sorted into 96-well plates (Corning) pre-filled with 100µl of methylcellulose and water in the outer wells or plated at 500 cells per 6-cm dish with methylcellulose (Methocult, StemCell Technologies). After 14 days, colonies were counted and scored as BFU-E, CFU-M, CFU-GM and CFU-GEMM according to the manual for "Human

Colony-forming Unit (CFU) Assays Using MethoCult" from StemCell Technologies and prior expertise (50). For DNA extraction from 96-well plates, PBS was added to wells with colonies, and the contents were mixed and transferred to a U-bottomed 96-well plate. From 6-cm dishes, colonies were picked and transferred to PBS. Cells were pelleted by centrifugation at 300xg for 5min followed by a wash with PBS. Finally, cells were resuspended in 25 µl QuickExtract DNA Extraction Solution (Epicentre, Madison, WI, USA) and transferred to PCR plates, which were incubated at 65˚C for 10 min followed by 100˚C for 2 min. For *CCR5*, a 3-primer PCR was set up with a forward primer binding in the left homology arm, a forward primer binding in the insert, and a reverse primer binding in CCR5 outside the right homology arm CCR5_inside_LHA: 5'-GCACAGGGTGGAACAAGATGG-3', CCR5_insert: 5'-AAGGGGGAGGATTGGGAAGAC-3', CCR5_outside_RHA: 5'-TCAAGAATCAGCAATTCTCTGAGGC-3'. For all other genes, gene-specific integration was detected by 'In-Out' PCR using a primer that binds outside the homology arm (HA) and a primer specific for the transgene cassette (insert). *HBB*_outside_LHA: GAAGATATGCTTAGAACCGAGG, *HBB*_insert: ACCGCAGATATCCTGTTTGG *IL2RG*_insert: 5'-GTACCAGCACGCCTTCAAGACC-3', *IL2RG*_outside_RHA: 5'-CAGATATCCAGAGCCTAGCCTCATC-3', *RUNX1*_outside_RHA: 5'-GAAGGGCATTGCTCAGAAAA-3', *RUNX1*_insert: 5'- AAGGGGGAGGATTGGGAAGAC-3', *ASXL1*_outside_RHA: 5'- AAGGGGGAGGATTGGGAAGAC-3', *ASXL1*_insert: 5'-CCTCCCAAGCTGGAACTACA-3'. For detecting IL2RG non-integrated (non_int) alleles the following primers were used: IL2RG_non_int_fw: 5'-TCACACAGCACATATTTGCCACACCCTCTG-3′, IL2RG_non_int_rv: 5′-TGCCCACATGATTGTAATGGCCAGTGG-3'. For detecting dual integration of GFP and tdTomato into two *HBB* alleles, a primer in *HBB* outside the right homology arm was used together with either a GFP or tdTomato-specific primer: *HBB*_outside_RHA: 5'-GATCCTGAGACTTCCACACTGATGC-3', GFP: 5'-GTACCAGCACGCCTTCAAGACC-3', tdTomato: 5'-CGGCATGGACGAGCTGTACAAG-3'. Clones with di-genic GFP (*HBB*)/mCherry

(*CCR5*) and tri-genic GFP (*IL2RG*)/tdTomato (*HBB*)/tNGFR (*CCR5*) integrations were screened for integrations using the same primers as above. All integrated PCR bands were subjected to Sanger sequencing to confirm perfect HR at the intended locus. For flow-cytometric analysis of colonies generated from cells with quadruple-allelic HR, individual colonies were picked and directly resuspended in FACS buffer containing LIVE/DEAD staining solution (LIVE/DEAD® Fixable Near-IR Dead Cell Stain, Thermo). After 30min incubation (4ºC, dark) cells were washed in FACS buffer and subjected to analysis. Dead cells were excluded from analysis based on APC-Cy7 positivity.

Transplantation of CD34$^+$ HSPCs into NSG mice

6 to 8-week-old NOD scid gamma (NSG) mice were used (Jackson laboratory, Bar Harbor, ME USA). The experimental protocol was approved by Stanford University's Administrative Panel on Lab Animal Care (IACUC 25065). Four days after electroporation/transduction, different populations of live (DAPI-negative) targeted cells were sorted. Mock-treated cells were also sorted to control for the effect of the sorting procedure. Directly after sorting, cells were transplanted into one femur of sub-lethally irradiated mice (200 rad, 24 hours before transplant). Mice were randomly assigned to each experimental group and analyzed in a blinded fashion.

Assessment of human engraftment

12 weeks after transplantation, mice were sacrificed, mouse bone marrow (BM) was harvested from the transplanted femur by flushing. Non-specific antibody binding was blocked (10% vol/vol, TruStain FcX, BioLegend) and cells were stained (30min, 4°C, dark) with monoclonal anti-human HLA-ABC APC-Cy7 (W6/32, BioLegend), anti-mouse CD45.1 PE-Cy7 (A20, eBioScience, San Diego, CA, USA), CD19 APC (HIB19, BD511 Biosciences), CD33 PE (WM53, BD Biosciences), and anti-mouse mTer119 PE-Cy5 (TER-119, BD Biosciences) antibodies, and Propidium Iodide to detect dead cells. Human engraftment was defined as HLA-ABC$^+$ cells.

Analysis of HBB-AAVS1 translocations

Genomic DNA was extracted from sorted populations using QuickExtract DNA Extraction Solution. For ddPCR quantification of translocations, ddPCR droplets were generated on a QX200 Droplet Generator (Bio-Rad) according to manufacturer's protocol. Briefly, PCR reactions were set up in a 25µL total volume per reaction with the ddPCR Supermix for Probes (No dUTP) (Bio-Rad). A HEX reference assay detecting copy number input of the *TERT* gene was used to normalize for genomic DNA input (Bio-Rad: saCP1000100). A custom assay designed to detect the translocations between *HBB* and *AAVS1* consisted of: Forward primer: 5'-TCAGGGCAGAGCCATCTATTGC-3', Reverse primer: 5'-CCAGATAAGGAATCTGCCTAACAGG-3', 5'-6FAM/ZEN/3'-IBFQ-labeled Probe (IDT): 5'-CTTCTGACACAACTGTGTTCACTAGCAACC-3'. The translocation assay was used at a final concentration of 900nM for each of the primers and a final concentration of 250nM for the probe. 20µL of the PCR reaction was used for droplet generation, and 40µL of the droplets was used in the following PCR conditions: 95° - 10 min, 50 cycles of 94° - 30s, 57°C – 30s, and 72° - 2min, finalize with 98° - 10min and 4°C until droplet analysis. Droplets were analyzed on a QX200 Droplet Reader (Bio-Rad) detecting FAM and HEX positive droplets. Control samples with non-template control (H_2O) or genomic DNA from mock-electroporated samples were included in the entire process. Translocation frequencies were calculated as the translocation copy number per µL divided by the TERT copy number per µL. For sequencing of translocations, PCR products were generated using Phusion polymerase (Fisher Scientific) with the forward and reverse primers listed above for the translocation ddPCR assay. PCR amplicons were gel-purified and cloned into the pMiniT 2.0 plasmid using the NEB PCR Cloning Kit (NEB) according to manufacturer's recommendations. Ligated plasmid reactions were transformed into XL-1 Blue competent cells, plated on ampicillin-containing agar plates, and single colonies were sequenced

by MCLAB (South San Francisco, CA, USA) using rolling circle amplification followed by sequencing using the following primer: 5'-ACCTGCCAACCAAAGCGAGAAC-3'.

Analysis of cell viability and proliferation

Modified cells were FACS-sorted into individual wells of a 96-well U bottom plate and expanded in HSPC retention media (see above) at a density of <100,000 cells per mL. To check viability and proliferation after multiplexed HR, cells from a single well were recovered and a known number of absolute counting beads (CountBright beads, Invitrogen) was added. Cells were stained with Ghost Dye Red 780 (Tonbo Biosciences) for 30min at 4ºC in the dark and analyzed on a FACS-Aria II without further manipulation to reduce potential cells loss. Viable cells were determined as GhostDye Red 780 negative and exact cell counts were assessed through concomitant acquisition of 10,000 beads. Cell counts were calculated based on ratio of beads to cells within the suspension.

ACKNOWLEDGMENTS

R.O.B. was supported through an Individual Postdoctoral grant (DFF–1333-00106B) and a Sapere Aude, Research Talent grant (DFF–1331-00735B) both from the Danish Council for Independent Research, Medical Sciences. D.P.D was supported through the Stanford Child Health Research Institute (CHRI) Grant and Postdoctoral Award. A.R. was supported by an Erwin Schroedinger Fellowship from the Austrian Research Council (FWF). M.H.P. gratefully acknowledges the support of the Amon Carter Foundation, the Laurie Kraus Lacob Faculty Scholar Award in Pediatric Translational Research and NIH grant support R01-AI097320, and R01-AI120766. R.M. gratefully acknowledges the support of the Stanford Ludwig Center for Cancer Stem Cell Research, the Stanford Child Health Research Institute (CHRI), and NIH grant support R01-CA188055. R.M. is a New York Stem Cell Foundation Robertson Investigator and Leukemia and Lymphoma Society Scholar. We thank David Russell (University of Washington) for the pDGM6 plasmid, the Binn's Program for Cord Blood Research (Stanford University) for cord blood-derived $CD34^+$ HSPCs, Sruthi Mantri (Stanford University) for isolation of $CD34^+$ HSPCs from cord blood, and Carmencita Nicolas for help with *in vivo* experiments. We also thank members of the Porteus and Majeti labs, for helpful input, comments and discussion.

COMPETING INTERESTS

Matthew Porteus has equity and consults for CRISPR Therapeutics. Ravindra Majeti has equity and consults for Forty Seven Inc.

REFERENCES:

1. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. Cell Stem Cell. 2012;10(2):120-36.
2. Chan SM, Thomas D, Corces-Zimmerman MR, Xavy S, Rastogi S, Hong WJ, et al. Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia. Nature medicine. 2015;21(2):178-84.
3. Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM. Gene therapy: therapeutic gene causing lymphoma. Nature. 2006;440(7088):1123.
4. Naldini L. Gene therapy returns to centre stage. Nature. 2015;526(7573):351-60.
5. Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nat Biotechnol. 2015;33(9):985-9.
6. Holt N, Wang J, Kim K, Friedman G, Wang X, Taupin V, et al. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat Biotechnol. 2010;28(8):839-47.
7. Saydaminova K, Ye X, Wang H, Richter M, Ho M, Chen H, et al. Efficient genome editing in hematopoietic stem cells with helper-dependent Ad5/35 vectors expressing site-specific endonucleases under microRNA regulation. Mol Ther Methods Clin Dev. 2015;1:14057.
8. Mandal PK, Ferreira LM, Collins R, Meissner TB, Boutwell CL, Friesen M, et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell. 2014;15(5):643-52.
9. Schumann K, Lin S, Boyer E, Simeonov DR, Subramaniam M, Gate RE, et al. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc Natl Acad Sci U S A. 2015;112(33):10437-42.
10. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014;24(6):1012-9.
11. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014;3:e04766.
12. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol. 2015;33(6):661-7.
13. Hultquist JF, Schumann K, Woo JM, Manganaro L, McGregor MJ, Doudna J, et al. A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. Cell Rep. 2016;17(5):1438-52.
14. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.
15. Heckl D, Kowalczyk MS, Yudovich D, Belizaire R, Puram RV, McConkey ME, et al. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. Nature biotechnology. 2014;32(9):941-6.
16. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell. 2014;159(2):440-55.
17. Brown A, Woods WS, Perez-Pinera P. Multiplexed Targeted Genome Engineering Using a Universal Nuclease-Assisted Vector Integration System. ACS Synth Biol. 2016;5(7):582-8.

564 18. Dever DP, Bak RO, Reinisch A, Camarena J, Washington G, Nicolas CE, et al.
565 CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. Nature.
566 2016;539(7629):384-9.
567 19. DeWitt MA, Magis W, Bray NL, Wang T, Berman JR, Urbinati F, et al. Selection-free
568 genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells.
569 Sci Transl Med. 2016;8(360):360ra134.
570 20. De Ravin SS, Li L, Wu X, Choi U, Allen C, Koontz S, et al. CRISPR-Cas9 gene repair of
571 hematopoietic stem cells from patients with X-linked chronic granulomatous disease. Sci
572 Transl Med. 2017;9(372).
573 21. Wang J, Exline CM, DeClercq JJ, Llewellyn GN, Hayward SB, Li PW, et al. Homology-
574 driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and
575 AAV6 donors. Nat Biotechnol. 2015;33(12):1256-63.
576 22. Hoban MD, Lumaquin D, Kuo CY, Romero Z, Long J, Ho M, et al. CRISPR/Cas9-
577 Mediated Correction of the Sickle Mutation in Human CD34+ cells. Mol Ther.
578 2016;24(9):1561-9.
579 23. Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, et al. Gene editing of CCR5 in
580 autologous CD4 T cells of persons infected with HIV. N Engl J Med. 2014;370(10):901-10.
581 24. Genovese P, Schiroli G, Escobar G, Di Tomaso T, Firrito C, Calabria A, et al. Targeted
582 genome editing in human repopulating haematopoietic stem cells. Nature.
583 2014;510(7504):235-40.
584 25. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, et al. Prognostic
585 relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med.
586 2012;366(12):1079-89.
587 26. Mazumdar C, Shen Y, Xavy S, Zhao F, Reinisch A, Li R, et al. Leukemia-Associated
588 Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic
589 Progenitor Differentiation. Cell Stem Cell. 2015;17(6):675-88.
590 27. Kotin RM, Linden RM, Berns KI. Characterization of a preferred site on human
591 chromosome 19q for integration of adeno-associated virus DNA by non-homologous
592 recombination. EMBO J. 1992;11(13):5071-8.
593 28. Viny AD, Ott CJ, Spitzer B, Rivas M, Meydan C, Papalexi E, et al. Dose-dependent role
594 of the cohesin complex in normal and malignant hematopoiesis. J Exp Med.
595 2015;212(11):1819-32.
596 29. Mullenders J, Aranda-Orgilles B, Lhoumaud P, Keller M, Pae J, Wang K, et al. Cohesin
597 loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative
598 neoplasms. J Exp Med. 2015;212(11):1833-50.
599 30. Seita J, Weissman IL. Hematopoietic stem cell: self-renewal versus differentiation.
600 Wiley interdisciplinary reviews Systems biology and medicine. 2010;2(6):640-53.
601 31. Maddalo D, Manchado E, Concepcion CP, Bonetti C, Vidigal JA, Han YC, et al. In vivo
602 engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system.
603 Nature. 2014;516(7531):423-7.
604 32. Poirot L, Philip B, Schiffer-Mannioui C, Le Clerre D, Chion-Sotinel I, Derniame S, et al.
605 Multiplex Genome-Edited T-cell Manufacturing Platform for "Off-the-Shelf" Adoptive T-cell
606 Immunotherapies. Cancer Res. 2015;75(18):3853-64.
607 33. Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T, et al. Uncovering
608 regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'.
609 Nature genetics. 2005;37(3):225-32.

610 34. Miller DG, Petek LM, Russell DW. Adeno-associated virus vectors integrate at
611 chromosome breakage sites. Nature genetics. 2004;36(7):767-73.
612 35. Fares I, Chagraoui J, Gareau Y, Gingras S, Ruel R, Mayotte N, et al. Cord blood
613 expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-
614 renewal. Science. 2014;345(6203):1509-12.
615 36. Cutler C, Multani P, Robbins D, Kim HT, Le T, Hoggatt J, et al. Prostaglandin-
616 modulated umbilical cord blood hematopoietic stem cell transplantation. Blood.
617 2013;122(17):3074-81.
618 37. de Lima M, McNiece I, Robinson SN, Munsell M, Eapen M, Horowitz M, et al. Cord-
619 blood engraftment with ex vivo mesenchymal-cell coculture. N Engl J Med.
620 2012;367(24):2305-15.
621 38. Popat U, Mehta RS, Rezvani K, Fox P, Kondo K, Marin D, et al. Enforced fucosylation of
622 cord blood hematopoietic cells accelerates neutrophil and platelet engraftment after
623 transplantation. Blood. 2015;125(19):2885-92.
624 39. Reinisch A, Thomas D, Corces MR, Zhang X, Gratzinger D, Hong WJ, et al. A humanized
625 bone marrow ossicle xenotransplantation model enables improved engraftment of healthy
626 and leukemic human hematopoietic cells. Nat Med. 2016;22(7):812-21.
627 40. Hubbard N, Hagin D, Sommer K, Song Y, Khan I, Clough C, et al. Targeted gene editing
628 restores regulated CD40L function in X-linked hyper-IgM syndrome. Blood.
629 2016;127(21):2513-22.
630 41. Voit RA, Hendel A, Pruett-Miller SM, Porteus MH. Nuclease-mediated gene editing by
631 homologous recombination of the human globin locus. Nucleic Acids Res. 2014;42(2):1365-
632 78.
633 42. Bak RO, Porteus MH. CRISPR-Mediated Integration of Large Gene Cassettes Using
634 AAV Donor Vectors. Cell Rep. 2017;20(3):750-6.
635 43. Zhao Z, Zuber J, Diaz-Flores E, Lintault L, Kogan SC, Shannon K, et al. p53 loss
636 promotes acute myeloid leukemia by enabling aberrant self-renewal. Genes & development.
637 2010;24(13):1389-402.
638 44. Sweeney CL, Zou J, Choi U, Merling RK, Liu A, Bodansky A, et al. Targeted Repair of
639 CYBB in X-CGD iPSCs Requires Retention of Intronic Sequences for Expression and
640 Functional Correction. Mol Ther. 2017;25(2):321-30.
641 45. Voit RA, McMahon MA, Sawyer SL, Porteus MH. Generation of an HIV resistant T-cell
642 line by targeted "stacking" of restriction factors. Mol Ther. 2013;21(4):786-95.
643 46. Eyquem J, Mansilla-Soto J, Giavridis T, van der Stegen SJ, Hamieh M, Cunanan KM, et
644 al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. Nature.
645 2017;543(7643):113-7.
646 47. Khan IF, Hirata RK, Russell DW. AAV-mediated gene targeting methods for human
647 cells. Nature protocols. 2011;6(4):482-501.
648 48. Aurnhammer C, Haase M, Muether N, Hausl M, Rauschhuber C, Huber I, et al.
649 Universal real-time PCR for the detection and quantification of adeno-associated virus
650 serotype 2-derived inverted terminal repeat sequences. Human gene therapy methods.
651 2012;23(1):18-28.
652 49. Bonini C, Grez M, Traversari C, Ciceri F, Marktel S, Ferrari G, et al. Safety of retroviral
653 gene marking with a truncated NGF receptor. Nature medicine. 2003;9(4):367-9.
654 50. Majeti R, Park CY, Weissman IL. Identification of a hierarchy of multipotent
655 hematopoietic progenitors in human cord blood. Cell Stem Cell. 2007;1(6):635-45.
656

FIGURE LEGENDS

Figure 1: FACS-based identification and enrichment of monogenic genome-edited CD34$^+$ human hematopoietic stem and progenitor cells (HSPCs). (a) HSPCs were electroporated with *CCR5*-RNP and transduced with *CCR5*-tNGFR rAAV6 HR donor. Representative FACS plots from day 4 post-electroporation highlight the *CCR5* tNGFR^{high} population (red gate) generated by the addition of Cas9 RNP compared to cells with low reporter expression (green gate) and reporter^{negative} cells (black gate). Numbers reflect percentage of cells within gates. **(b)** Day 4 post-electroporation, *CCR5* (tNGFR or GFP) and *IL2RG* (GFP)-targeted HSPCs from reporter^{high} (red), reporter^{low} (green), and reporter^{neg} (blue) fractions were sorted and cultured for 21 days while monitoring the percentage of cells that remained GFP$^+$. Error bars represent S.E.M. *N* = 6 for *CCR5*, *N* = 3 for *IL2RG*, all from different CD34$^+$ donors. **(c)** HSPCs were targeted at *CCR5* (with GFP or tNGFR donor) or at *IL2RG* (GFP donor; only female cells for *IL2RG*). At day 4 post-electroporation, reporter^{high} cells were single-cell sorted into methylcellulose for colony formation. PCR was performed on colony-derived gDNA to detect targeted integrations. 338 *CCR5* and 177 *IL2RG* myeloid and erythroid methylcellulose colonies were screened from at least two different CD34$^+$ HSPC donors. **(d)** HSPCs were targeted at the *STAG2* gene or the *AAVS1* locus with a GFP reporter cassette. Cells that only received the *STAG* 2 GFP AAV6 donor and not Cas9 RNP were included as an additional control. At day 4 post-electroporation and transduction, reporter^{high} cells from the *STAG2* and *AAVS1* targeting experiments and bulk cells from the *STAG2* AAV6 only population were plated in methylcellulose for colony formation. After 14 days, colonies were scored as either erythroid or myeloid based on morphology. Error bars represent S.E.M, N = 3, *** $p < 0.001$, n.s. = $p \geq 0.05$, unpaired t-test.

Figure 2: Identification and enrichment of biallelic genome-edited CD34$^+$ human hematopoietic stem and progenitor cells (HSPCs). (a) *Left,* Schematic showing biallelic targeting strategy for *ASXL1* using GFP and BFP-encoding rAAV6 donors for integration into

each allele of *ASXL1*. The SFFV promoter drives reporter expression. *Middle, FACS plot from an* 'AAV only' sample day 4 post electroporation, showing low episomal reporter expression (BFP and GFP) in cells without the CRISPR system. *Right*, FACS plot of $CD34^+$ HSPCs treated with both Cas9 RNP and the two rAAV6 donors highlighting the generation of BFP^{high}/GFP^{high} double positive cells that have undergone *ASXL1* dual-allelic targeting. **(b)** HSPCs were targeted at both alleles of *HBB* (Cas9 RNP with GFP and tdTomato rAAV6 donors) and at day 4 post electroporation, dual positive cells were sorted and cultured for 16 days while analyzing reporter expression. Error bars representing S.E.M. are present, but too small to be visible (*N* = 3 different HSPC donors). **(c)** Gel images showing PCR genotyping of six methylcellulose-derived clones from (e) confirming integration into each of the *HBB* alleles. **(d)** Human primary T cells were CD3/CD28 stimulated for three days and then electroporated with *CCR5*-targeting Cas9 RNP and transduced with two *CCR5*-specific rAAV6 donors encoding GFP and mCherry, respectively. FACS plots show $GFP^{high}/mCherry^{high}$ biallelic targeting frequencies at day 4 post-electroporation.

Figure 3: Identification, enrichment, and long-term engraftment in NSG mice of di-genic genome-edited $CD34^+$ human hematopoietic stem and progenitor cells (HSPCs). (a) *Left,* Schematic depicting *HBB* and *IL2RG* di-genic targeting. *Middle,* FACS plot of an 'AAV only' sample at day 4 post electroporation, showing low episomal reporter expression (*HBB*-tdTomato and *IL2RG*-GFP) in cells without the CRISPR system. *Right*, FACS plot at day 4 post-electroporation of HSPCs electroporated with Cas9 RNP targeting both *HBB* and *IL2RG* followed by transduction with *HBB*-tdTomato and *IL2RG*-GFP rAAV6 donors showing the generation of $tdTomato^{high}/GFP^{high}$ cells with di-genic targeting at *HBB* and *IL2RG*. **(b)** Double-positive HSPCs targeted at *HBB* (GFP) and *CCR5* (mCherry) were sorted at day 4 post-electroporation and cultured for 15 days while analyzing reporter expression. Error bars represent S.E.M. (*N* = 3 different HSPC donors). **(c)** Representative gel images showing PCR genotyping of six (out of 57

total) *HBB*-GFP^{high} (gene reporter 1)/*CCR5*-mCherry^{high} (gene reporter 2) methylcellulose-derived clones confirming integration at each locus **(d)** Representative fluorescence microscopy images of methylcellulose-derived clones with di-genic targeting at *HBB* and *CCR5* show myeloid and erythroid progenitors with both GFP and mCherry expression. **(e)** HSPCs were targeted at the *HBB* and *AAVS1* loci with a GFP and BFP expression cassette, respectively. Representative FACS plot (left panel) shows analysis seven days after targeting. All four gated populations were sorted and genomic DNA was subject to TIDE analysis for determining INDEL frequencies at the two loci (middle panel), and subject to ddPCR quantification of one of the two possible monocentric translocations between *HBB* and *AAVS1* (right panel) (see also Figure 3–figure supplement 2). **(f)** Representative FACS plots from cells targeted at the *HBB* and *AAVS1* loci with a GFP and BFP expression cassette, respectively. Representative FACS plot shows analysis four days after targeting at which point the four populations were sorted and transplanted intrafemorally into NSG mice that were irradiated 24 hrs before transplantation. **(g)** Bone marrow from the injected femurs from the mice transplanted as described in (f) was analyzed 12 weeks after transplantation. Representative FACS plots are from a mouse from each of the four groups depicted in (f) as well as a mouse transplanted with mock-electroporated cells. The middle row depicts human engraftment gated as positive for the human leukocyte antigen complex (HLA-ABC). The upper and lower rows depict FACS plots gated from the human populations and show myeloid (CD33^{+}) and lymphoid (CD19^{+}) engraftment (upper row) as well as reporter gene expression (lower row) (see also Figure 3-figure supplement 3 for all transplantation data).

Figure 4: Multiplexing homologous recombination in CD34^{+} human hematopoietic stem and progenitor cells (HSPCs). (a) HSPCs were electroporated with Cas9 RNP targeting *ASXL1* and *RUNX1* followed by rAAV6 transduction with two donors for *ASXL1* (mCherry and GFP) and two donors for *RUNX1* (E2Crimson and BFP). Tetra-allelically targeted HSPCs were identified as

mCherry^{high}/GFP^{high}/BFP^{high}/E2Crimson^{high} (*N* = 3 see **Supplementary file 1e**) **(b)** Cells modified at both alleles for *RUNX1* and *ASXL1* (as in (a)) were subjected to a methylcellulose assay (triplicates) and scored as BFU-E, CFU-M, CFU-GM or CFU-GEMM based on morphology 14 days after sorting. **(c)** PCR was performed on colony-derived gDNA to detect targeted integrations at both genes. 73 individual colonies were analyzed. Color coding for colonies with triple-allelic integration are as follows: grey: *RUNX1* biallelic/*ASXL* monoallelic; white: *RUNX1* monoallelic/*ASXL1* biallelic. **(d)** For tri-genic targeting of HSPCs, cells were electroporated with Cas9 RNP targeting *IL2RG*, *HBB*, and *CCR5* followed by transduction of three rAAV6 donors homologous to each of the three genes (*IL2RG*-GFP, *HBB*-tdTomato, and *CCR5*-tNGFR). Tri-genic-targeted cells were identified as reporter^{high} for all three reporters (*N* = 5 see **Supplementary file 1e**). **(e)** Methylcellulose clones from the triple-positive cells in (d) were subjected to genotyping PCR and gel images show colonies with targeted integration at all three genes in 9/11 colonies (note that GFP shows a faint band in colony 6). **(f)** *Left*, Schematic showing strategy for targeting four different genes (*HBB*, *RUNX1*, *ASXL1*, and *CCR5*) simultaneously (tetra-genic). Four different genes are targeted by electroporation of four different Cas9 RNPs followed by transduction with four different rAAV6 donors that each targets a gene with a different reporter. *Right*, Tetra-genic targeting at the above-mentioned four genes was identified as reporter^{high} for all four reporters (*N* = 3 see **Supplementary file 1e**).

Figure 1-figure supplement 1. Analysis of cell fractions with different fluorescence intensity. a) Schematic showing the general layout of the AAV6 donor employed. ITR: inverted terminal repeat; SFFV promoter: spleen focus forming virus promoter; GFP: green fluorescent protein; polyA: bovine growth hormone polyadenylation signal; RHA: right homology arm. Approximate sizes are shown below each component. **b)** Cells were targeted at the HBB locus by electroporation of Cas9 RNP followed by transduction of a homologous rAAV6 donor carrying a GFP expression cassette. At 4 days post electroporation and transduction, cells with different

762 GFP intensities (GFP^{high}, GFPLow^{High}, GFPLow^{Med}, GFPLow^{Low}) were FACS-sorted and cultured
763 for an additional 16 days. At day 20 post targeted, cells were analyzed for GFP expression by
764 flow cytometry and the red gates show the GFP^{high} gate at this time point. **c)** The cells from a)
765 were analyzed at different time points after sorting, and data points show the percentage of cells
766 within the GFP^{high} gate for the different populations as well as a population receiving only the
767 rAAV6 donor and not Cas9 RNP.
768
769 **Figure 1-figure supplement 2. Genotypes of clones with mono-genic targeting. a)** *Left*,
770 schematic representation of the three-primer PCR used to genotype *CCR5* alleles for integrated
771 (red PCR product) and non-integrated (green PCR product) alleles. One forward primer is
772 located in the left homology arm (LHA), one forward primer is located in the poly A, and a
773 common reverse primer is located outside the region of the right homology arm (RHA). *Right*, gel
774 image of representative genotyped clones from Figure 1c (*CCR5*) showing colonies with biallelic
775 and monoallelic integrations. **b)** A subset of the *CCR5* and *IL2RG* clones (only female cells for
776 *IL2RG*) from Figure 1c with monoallelic integration had the genotype on the non-integrated allele
777 analyzed by Sanger sequencing of purified PCR products. Note that in-frame INDELs can be
778 gene-disrupting depending on the location and size of the INDEL. **c)** As in Figure 1c, HSPCs
779 were targeted at *RUNX1* and at day 4 post-electroporation, reporter^{high} cells were single cell-
780 sorted into methylcellulose-containing 96-well plates to establish colonies. After 14 days, PCR
781 was performed on colony-derived gDNA to detect targeted integrations. A total of 36 myeloid and
782 erythroid methylcellulose colonies were screened. **d)** The monoallelically targeted clones from c)
783 had the genotype assessed on the non-integrated allele by Sanger sequencing of purified PCR
784 products. See Supplementary file 1b for complete list of genotypes.
785
786 **Figure 2-figure supplement 1. Cas9 and rAAV6-mediated biallelic homologous**
787 **recombination (HR) in human CD34^{+} HSPCs.** *Top*, Representative FACS plots from HSPCs

788 transduced with two rAAV6 (two fluorescent reporters for each gene) that have homology for the
789 genes listed on the bottom panel show low episomal expression and very few dual reporter^{high}
790 expressing HSPCs. *Bottom*, HSPCs were electroporated with gene-specific Cas9 RNPs and
791 then transduced with rAAV6 targeting each allele of a gene with two different indicated
792 fluorescent reporters**.** FACS plots at Day 4-post electroporation highlight the dual reporter^{high}
793 cells that have undergone HR at both alleles of the intended gene.
794
795 **Figure 2-figure supplement 2. Toxicity assessment of biallelic integration at the *CCR5***
796 **locus in primary human T cells.** Human primary T cells were isolated from buffy coats and
797 stimulated for three days using anti-CD3 and anti-CD28 antibodies. Cells were then
798 electroporated with *CCR5*-targeting Cas9 RNP and transduced with two *CCR5*-specific rAAV6
799 donors encoding GFP and mCherry, respectively, either alone or in combination. Cell viabilities
800 were measured at Day 2 post-electroporation by Trypan Blue exclusion assay (*N* = 2 different
801 buffy coat-derived T cells).
802
803 **Figure 3-figure supplement 1. Cas9 and rAAV6-mediated di-genic homologous**
804 **recombination (HR) in human CD34^{+} HSPCs.** *Top*, Representative FACS plots of HSPCs
805 transduced with two rAAV6 donors targeting two genes with two distinct fluorescent reporters
806 (listed in FACS plots in lower panel) show low episomal expression and few dual reporter^{high}
807 expressing HSPCs. *Bottom*, HSPCs were electroporated with two different gene-specific Cas9
808 RNPs and then transduced with homologous rAAV6 donors (each gene targeted with a different
809 fluorescent reporter). Representative FACS plots from Day 4-post electroporation show the
810 generation of dual reporter^{high} positive HSPCs targeted at both genes.
811
812 **Figure 3-figure supplement 2. Measuring translocations after *HBB* and *AAVS1* di-genic**
813 **targeting. (a)** Schematic showing the *HBB* gene on chromosome 11 and the *AAVS1* locus on

chromosome 19. The Cas9 cut sites are shown in red. One of the two possible monocentric translocations is shown. **(b)** The reference sequence of the *HBB-AAVS1* translocation is shown in the top. Below are representative translocation sequences from $GFP^{-}BFP^{-}$ HSPCs sorted seven days after targeting (see Figure 3e, left panel). **(c)** Representative ddPCR analyses quantifying translocations in NTC (non-template control), mock-electroporated, and $GFP^{-}BFP^{-}$ cells (see Figure 3e, right panel). The reference assay quantifies TERT gene copies used to normalize for DNA input. The translocation assay probe binds 50bp away from the junction and none of the identified translocations would therefore exclude probe binding.

Figure 3-figure supplement 3. Analysis of mice transplanted with different sorted populations of cells targeted at the *HBB* and *AAVS1* locus. The table shows an overview of the 11 NSG mice that were transplanted intrafemorally with either mock-electroporated cells or sorted cells from the four populations displayed in Fig. 3f. 12 weeks after transplant, the transplanted femurs were flushed and the cells analyzed for human engraftment based on HLA-ABC expression, B cell or myeloid phenotype (CD19 and CD33, respectively), and expression of the two reporter genes.

Figure 4-figure supplement 1. Targeting two genes for biallelic homologous recombination (HR) in primary $CD34^{+}$ HSPCs. (a) Schematic showing experimental strategy for Figure 4a for targeting both alleles of *RUNX1* and *ASXL1*. **(b)** FACS plots, gating scheme, and frequencies of HR at each allele for the experiment shown in Figure 4a. **(c)** FACS plot showing very low frequency of tetra-$reporter^{high}$ cells without Cas9. **(d)** FACS plots of cells from single methylcellulose colonies derived from tetra-$reporter^{high}$ cells from Figure 4a. **(e)** Schematic showing targeting both alleles of *RUNX1* and *HBB* for HR with four distinct reporters. **(f)** *Top*, FACS plots of HSPCs transduced with four rAAV6s (no Cas9 RNPs) showing the gating scheme and low episomal reporter expression without a nuclease. *Bottom*, HSPCs were electroporated

840 with RNPs targeting *HBB* and *RUNX1* and then transduced with four rAAV6s. FACS plots from
841 day 4 post electroporation show MFI shift for each reporter alone. *HBB*-tNGFR rAAV6 has
842 reproducibly shown lower episomal expression than all other rAAV6 we have used. **(g)** Images
843 from fluorescence microscopy showing an mCherry/BFP/GFP positive CFU-GM clone that has
844 undergone tetra-allelic HR. The colony was not stained for *HBB*-tNGFR. **(h)** *Left*, FACS plots
845 show very low frequency of tetra-reporter^{high} cells without Cas9. *Right*, Nuclease addition
846 increases the frequency of bi and tetra-reporter^{high} HSPCs.
847
848 **Figure 4-figure supplement 2. Multiplexing homologous recombination at three genes**
849 **simultaneously in HSPCs. (a)** Schematic showing experimental strategy for Figure 4d
850 targeting three genes, *IL2RG*, *CCR5*, and *HBB*. **(b)** FACS plots show gating scheme and HR
851 frequencies at each locus for the experiment shown in Figure 4d. **(c)** Schematic outlining another
852 tri-genic targeting experiment for *RUNX1*, *ASXL1,* and *HBB*. **(d)** *Top*, FACS plots of HSPCs
853 transduced with three rAAV6s (no RNPs). *Bottom*, HSPCs were electroporated with gene-
854 specific RNPs and then transduced with three rAAV6s. FACS plots at Day 4-post electroporation
855 show MFI shift for each reporter alone. **(e)** FACS plots from same sample as in (d), but showing
856 different combinations of di-genic reporter^{high} populations that contain the same frequency of tri-
857 genic reporter^{high} cells.
858
859 **Figure 4-figure supplement 3. Toxicity assessment of multiplexed HR.** CD34^{+} cells from
860 mobilized peripheral blood were targeted at one, two, or three genes with Cas9 RNP and rAAV6
861 donors. Viabilities were measured by flow cytometry 72hrs post-electroporation using Live/Dead
862 and Annexin V stains. Viable cells are defined as live, non-apoptotic (Annexin V^{-}) and plotted as
863 percentage of a single AAV6 donor alone. Error bars represent SD, ns = not statistically
864 significant, Mann-Whitney test, *N* = 2 different HSPC donors.
865

Figure 4-figure supplement 4. Assessment of false-positive frequencies of FACS-based identification of multiplexed HR in HSPCs. Since capture of rAAV6 donors at the site of a DSB via NHEJ has been reported, we measured the false-positive rate of multiplexing HR via flow cytometry. **(a)** False-positive frequencies of di-genic targeting in HSPCs was determined by electroporating cells with an *HBB*-targeting Cas9 RNP followed by transduction with *HBB*-GFP (homologous) and *CCR5*-mCherry (non-homologous) rAAV6 donors. FACS plots show a false-positive rate of 0.24% dual reporter^{high} cells. Note that 4% dual reporter^{high} cells was reported in Figure 3-figure supplement 1 when performing di-genic targeting at *CCR5* and *HBB*, giving a false positive rate of 6% of targeting. **(b)** *Left*, To determine false-positive frequencies of tri-genic targeting in HSPCs, we electroporated *IL2RG*-RNP and *HBB*-RNP into HSPCs followed by transduction with the rAAV6 donors *IL2RG*-GFP (homologous), *HBB*-tdTomato (homologous), and *CCR5*-tNGFR (non-homologous). FACS plots show a false-positive frequency of 0.47%. Note that Fig. 4d shows a tri-genic targeting frequency of 4.1% (a false-positive rate of 11% of targeting). *Right*, We employed a similar strategy to determine false-positive frequencies of tri-genic targeting, but this time used different combinations of on-target nucleases. The false-positive rate detected here was 0.1% (2.4% of targeting). **(c)** To determine tetra-genic false-positive frequencies, we electroporated HSPCs with three on-target nucleases (*IL2RG*, *HBB*, and *CCR5*) and then transduced with three homologous rAAV6 donors (*IL2RG*, *HBB*, and *CCR5*) and one non-homologous donor (*CXCL12*). FACS plots show a frequency of 0.09% that are reporter^{high} for all four reporters with Figure 4f showing a tetra-genic targeting frequency of 1.0% (a false-positive rate of 9% of targeting).

Figure 4-figure supplement 5. Controlling genotype with cDNA knock-in. (a) A heterozygous knockout population can be generated with two HR donors. The first donor is designed to knock-in a wild-type (WT) cDNA cassette into the start codon (ATG) of the gene of interest followed by a cassette encoding a reporter gene (here GFP). WT cDNA is expressed

892 from the endogenous promoter as reported by Voit *et al.*, (NAR, 2014), Hubbard *et al.*, (Blood,
893 2016), and Dever *et al.*, Nature, 2016), which maintains endogenous regulatory control over
894 gene expression. The other donor encodes another reporter (here BFP), which disrupts the
895 targeted gene. Double positive cells (GFP^+/BFP^+) are heterozygous for the knockout allele. **(b)** A
896 population heterozygous for a particular SNP can be generated using two donors that knock in
897 cDNA expression cassettes followed by different reporter genes. One cDNA is WT while the
898 other carries the SNP of interest. Double positive cells (GFP^+/BFP^+) are heterozygous for the
899 SNP allele. Endogenous 3' UTRs may be incorporated to preserve posttranscriptional regulation.
900 Heterozygous SNP cDNA knock-in may be expanded to two or more genes, which may be of
901 particular interest in studies of leukemia-mutated genes such as *DNMT3A*, *IDH1/2*, *JAK2*, and
902 *KRAS*, which often occur in various combinations as heterozygous gain-of-function or dominant
903 negative mutations. In addition, reporter knock-in combined with WT cDNA knock-in (as depicted
904 in a) as well as SNP cDNA knock-in (SNP that disrupts gene or gene function) combined with
905 WT cDNA knock-in (as depicted in b) could be used to study haploinsufficiencies. Though not
906 depicted, all genes in the schematic are followed by polyadenylation signals.
907
908 **Supplementary file 1a. Overview of Cas9 and rAAV6 mono-genic targeting experiments**
909 **performed in cord blood (CB), bone marrow (BM), and mobilized peripheral blood (mPB)-**
910 **derived human $CD34^+$ HSPCs**. This table summarizes all independent experiments targeting
911 *HBB*, *CCR5*, *IL2RG*, *RUNX1*, *ASXL1*, *STAG2*, and *AAVS1* in HSPCs and the reporter genes
912 used. GFP: green fluorescent protein, tNGFR: truncated Nerve Growth Factor Receptor, BFP:
913 blue fluorescent protein. Efficiencies were averaged across 47 independent experiments, $N = 47$.
914
915 **Supplementary file 1b. Overview of genotypes for the non-integrated alleles in mono-**
916 **genic integration experiments**. The three tables show the different INDELs that were identified
917 by Sanger Sequencing of the non-edited allele in mono-genic targeting experiments (*CCR5*,

IL2RG, and *RUNX1*) used to analyze genotype frequencies shown in Figure 1-figure supplement 2b and 2d. Alleles are grouped into WT (blue), INDELs that preserve the reading frame (red) and INDELs that disrupt the reading frame (green). Note that INDELs that preserve the reading frame can potentially be disruptive depending on the size and location. For example, the 147bp deletion in *RUNX1* is considered disruptive because of its large size and because it deletes the splice donor site in the intron between exon 2 and 3. For *IL2RG*, one clone was found to have an allele with integration of 230bp from the donor (at the end of RHA and 72bp into the ITR).

Supplementary file 1c. Overview of di-genic and biallelic targeting experiments in cord blood (CB), bone marrow (BM), and mobilized peripheral blood (mPB)-derived human CD34$^+$ HSPCs. This table summarizes the experiments targeting HSPCs for biallelic and di-genic HR and the reporter genes used. GFP: green fluorescent protein, tNGFR: truncated Nerve Growth Factor Receptor, BFP: blue fluorescent protein. Efficiencies were averaged across 16 and 17 independent experiments, respectively, $N = 16$ and $N = 17$.

Supplementary file 1d. Overview of genotypes for the non-integrated alleles in clones with tri-genic integrations. Each row of the table represents the genotype of a colony established from a tri-genic targeting experiment (*IL2RG*, *HBB*, and *CCR5*). Alleles are grouped into WT (blue), INDELs that preserve the reading frame (red) and INDELs that disrupt the reading frame (green). Note that INDELs that preserve the reading frame can potentially be disruptive depending on the size and location. For *HBB* we identified one clone where *HBD* had been used as repair template and three clones with mono-allelic integration of part of the SFFV promoter indicative of HR events that ended prematurely.

Supplementary file 1e. Overview of tetra-allelic, tri-genic, and tetra-genic targeting experiments performed in human CD34$^+$ HSPCs derived from cord blood (CB), bone

944 **marrow (BM), and mobilized peripheral blood (mPB)**. This table summarizes the independent
945 multiplexing HR experiments performed for tetra-allelic, tri-genic, and tetra-genic targeting and
946 the reporter genes used. GFP: green fluorescent protein, tNGFR: truncated Nerve Growth Factor
947 Receptor, BFP: blue fluorescent protein. Efficiencies were averaged across independent
948 experiments, *N* = 3 (tetra-allelic and tetra-genic) and *N* = 6 (tri-genic).
949

950 **Table 1: Overview of targeting experiments in hematopoietic stem and progenitor cells**
951 **(HSPCs).** Overview of all HSPC targeting experiments performed in this study with the number
952 of independent experiments (N) for each experiment type, and the mean targeting efficiency (±
953 SEM). See also supplementary file 1a, 1c, and 1e.

954 ***Table 1***

Monogenic	**47**	**21.7 ± 13.4**
Biallelic	**16**	**5.5 ± 4.2**
Di-genic	**17**	**8.1 ± 8.1**
Tetra-allelic	**3**	**0.9 ± 0.3**
Tri-genic	**6**	**4.5 ± 4.8**
Tetra-genic	**3**	**0.7 ± 0.3**

955

Figure 1

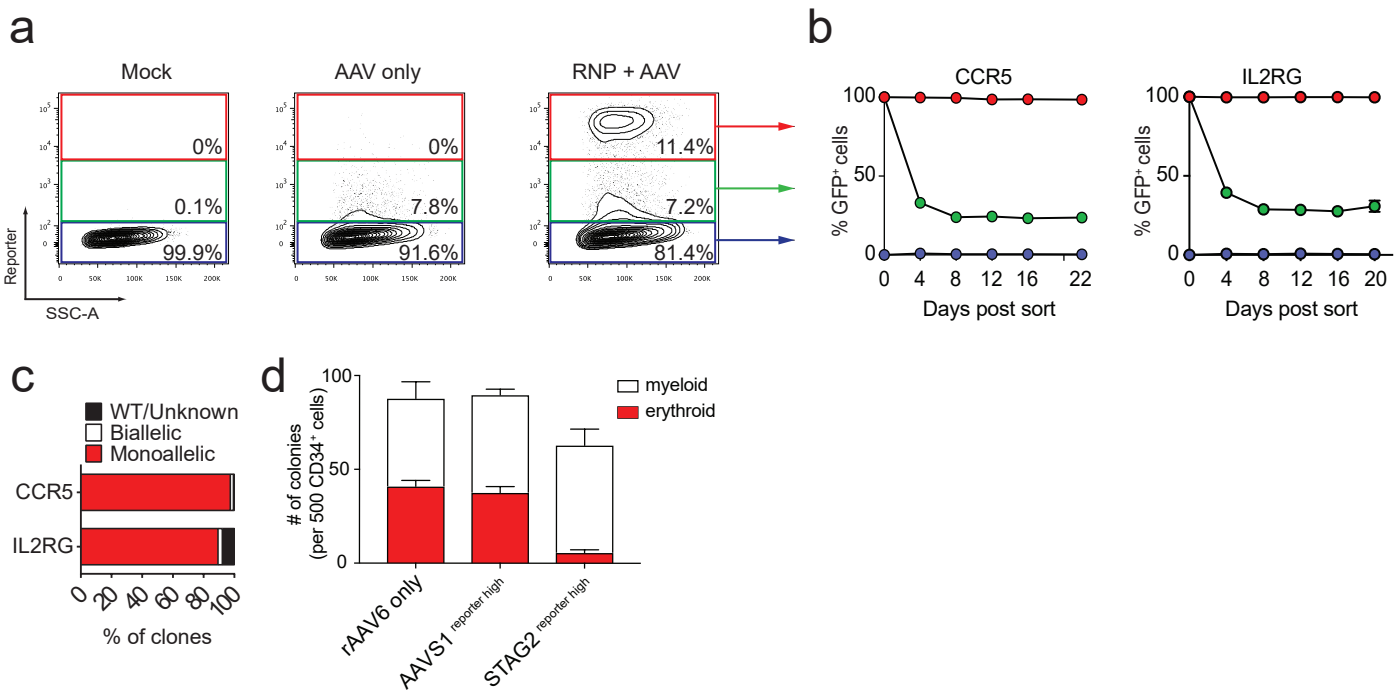

a
Mock
AAV only
RNP + AAV
0%
0.1%
99.9%
0%
7.8%
91.6%
11.4%
7.2%
81.4%
Reporter
SSC-A
b
CCR5
IL2RG
% GFP+ cells
Days post sort
c
WT/Unknown
Biallelic
Monoallelic
CCR5
IL2RG
% of clones
d
myeloid
erythroid
# of colonies (per 500 CD34+ cells)
rAAV6 only
AAVS1 reporter high
STAG2 reporter high

Figure 2

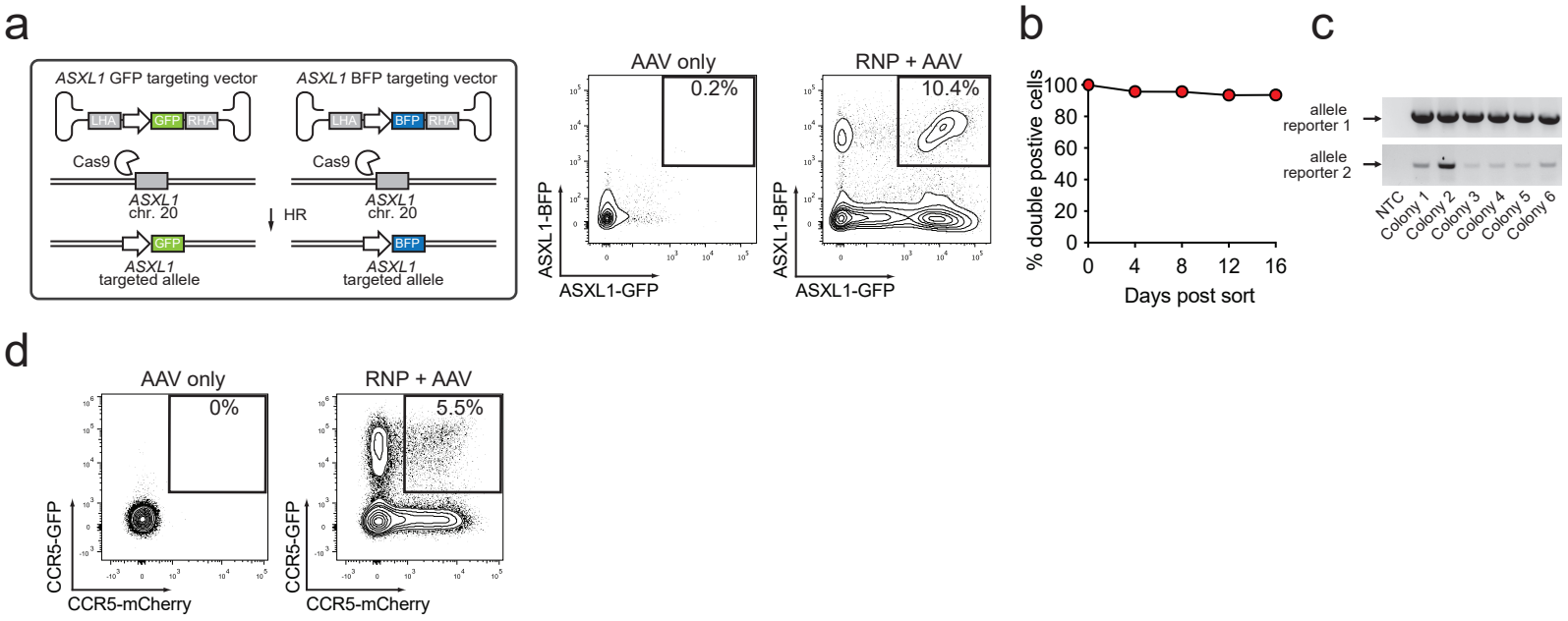

a
ASXL1 GFP targeting vector
ASXL1 BFP targeting vector
LHA
GFP
RHA
BFP
Cas9
ASXL1
chr. 20
HR
ASXL1
targeted allele
AAV only
0.2%
RNP + AAV
10.4%
ASXL1-BFP
ASXL1-GFP
b
% double postive cells
Days post sort
c
allele
reporter 1
allele
reporter 2
NTC
Colony 1
Colony 2
Colony 3
Colony 4
Colony 5
Colony 6
d
AAV only
0%
RNP + AAV
5.5%
CCR5-GFP
CCR5-mCherry

Figure 3

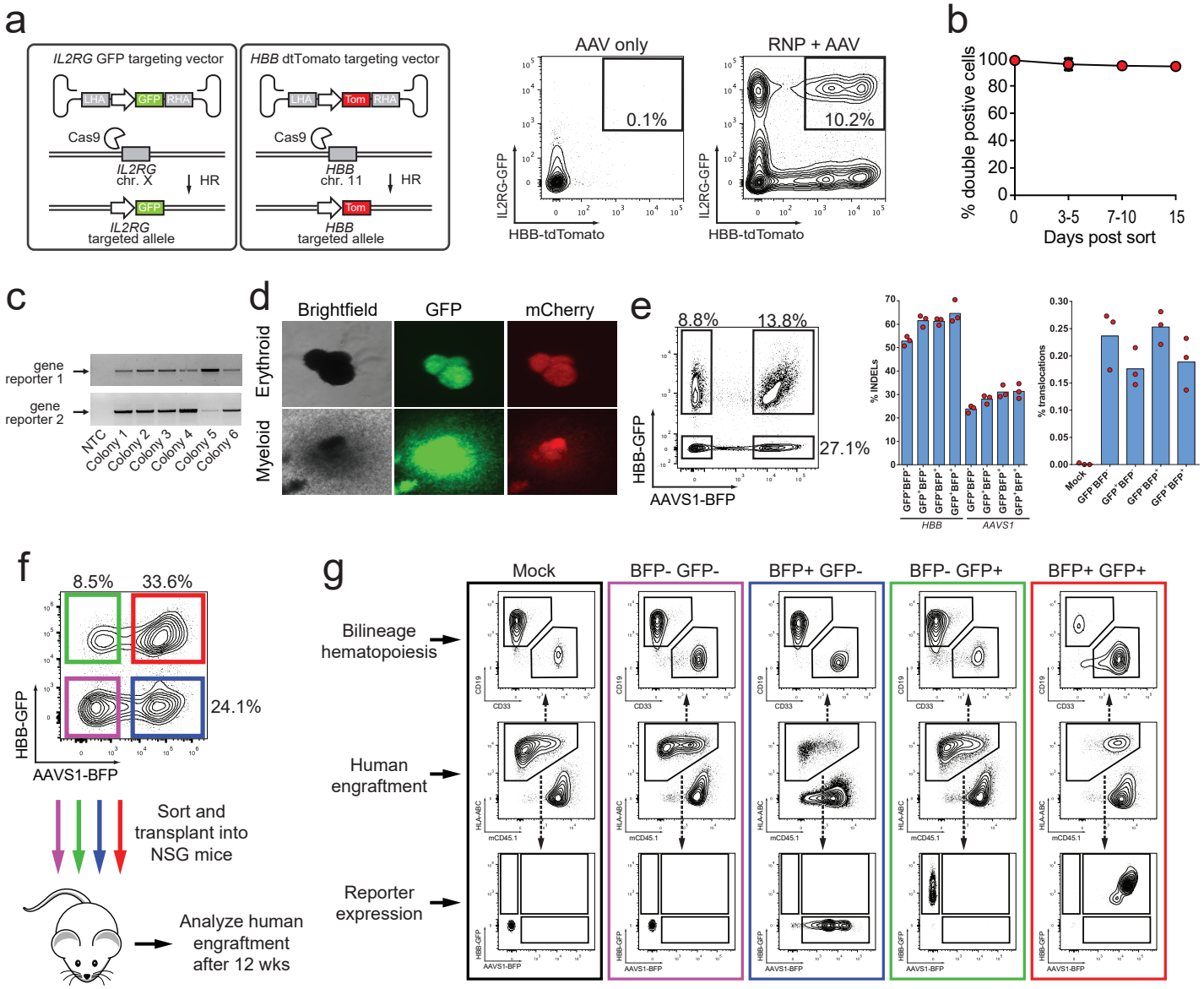
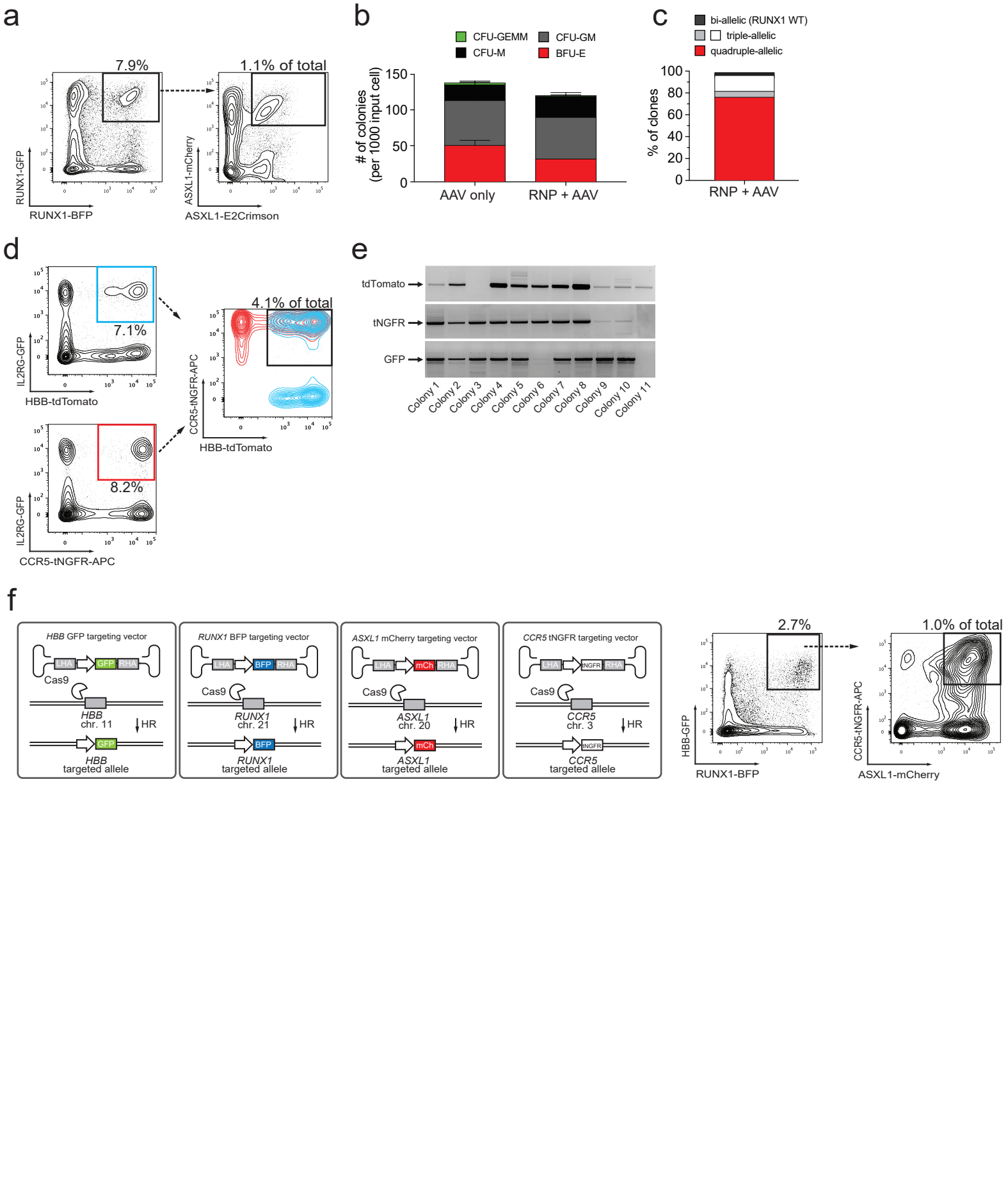

a
IL2RG GFP targeting vector
LHA
GFP
RHA
Cas9
IL2RG
chr. X
HR
IL2RG
targeted allele
HBB dtTomato targeting vector
Tom
HBB
chr. 11
HBB
targeted allele
AAV only
0.1%
RNP + AAV
10.2%
IL2RG-GFP
HBB-tdTomato
b
% double postive cells
Days post sort
c
gene reporter 1
gene reporter 2
NTC
Colony 1
Colony 2
Colony 3
Colony 4
Colony 5
Colony 6
d
Brightfield
GFP
mCherry
Erythroid
Myeloid
e
8.8%
13.8%
27.1%
HBB-GFP
AAVS1-BFP
% INDELs
% translocations
HBB
AAVS1
Mock
f
8.5%
33.6%
24.1%
HBB-GFP
AAVS1-BFP
Sort and transplant into NSG mice
Analyze human engraftment after 12 wks
g
Mock
BFP- GFP-
BFP+ GFP-
BFP- GFP+
BFP+ GFP+
Bilineage hematopoiesis
Human engraftment
Reporter expression
CD19
CD33
HLA-ABC
mCD45.1
HBB-GFP
AAVS1-BFP

Figure 4


a
7.9%
1.1% of total
RUNX1-GFP
RUNX1-BFP
ASXL1-mCherry
ASXL1-E2Crimson
b
CFU-GEMM
CFU-GM
CFU-M
BFU-E
# of colonies (per 1000 input cell)
150
100
50
0
AAV only
RNP + AAV
c
bi-allelic (RUNX1 WT)
triple-allelic
quadruple-allelic
% of clones
100
80
60
40
20
0
RNP + AAV
d
7.1%
IL2RG-GFP
HBB-tdTomato
4.1% of total
CCR5-tNGFR-APC
HBB-tdTomato
8.2%
IL2RG-GFP
CCR5-tNGFR-APC
e
tdTomato
tNGFR
GFP
Colony 1
Colony 2
Colony 3
Colony 4
Colony 5
Colony 6
Colony 7
Colony 8
Colony 9
Colony 10
Colony 11
f
HBB GFP targeting vector
LHA
GFP
RHA
Cas9
HBB
chr. 11
HR
GFP
HBB
targeted allele
RUNX1 BFP targeting vector
LHA
BFP
RHA
Cas9
RUNX1
chr. 21
HR
BFP
RUNX1
targeted allele
ASXL1 mCherry targeting vector
LHA
mCh
RHA
Cas9
ASXL1
chr. 20
HR
mCh
ASXL1
targeted allele
CCR5 tNGFR targeting vector
LHA
tNGFR
RHA
Cas9
CCR5
chr. 3
HR
tNGFR
CCR5
targeted allele
2.7%
1.0% of total
HBB-GFP
RUNX1-BFP
CCR5-tNGFR-APC
ASXL1-mCherry

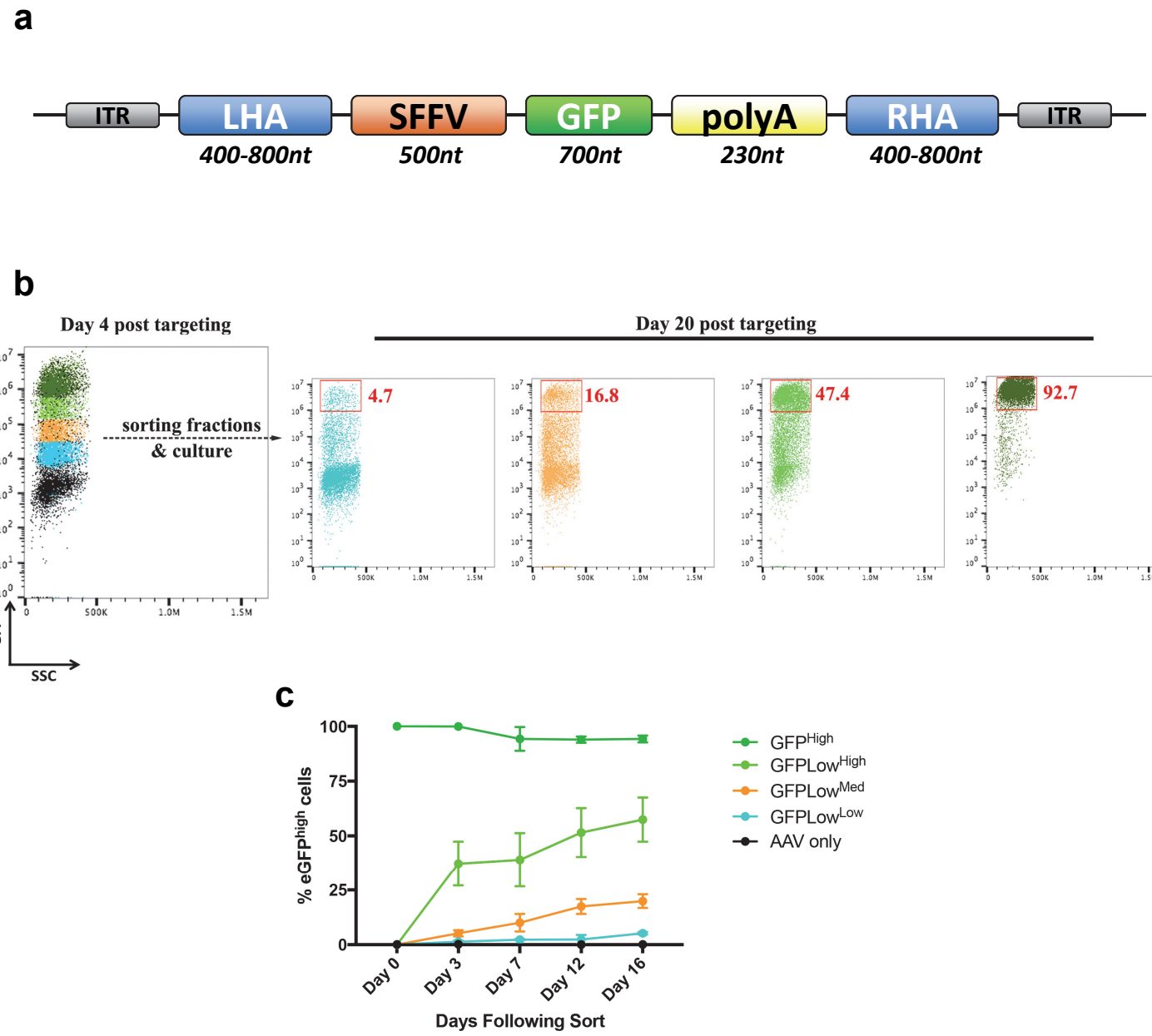


Figure 1-figure supplement 1

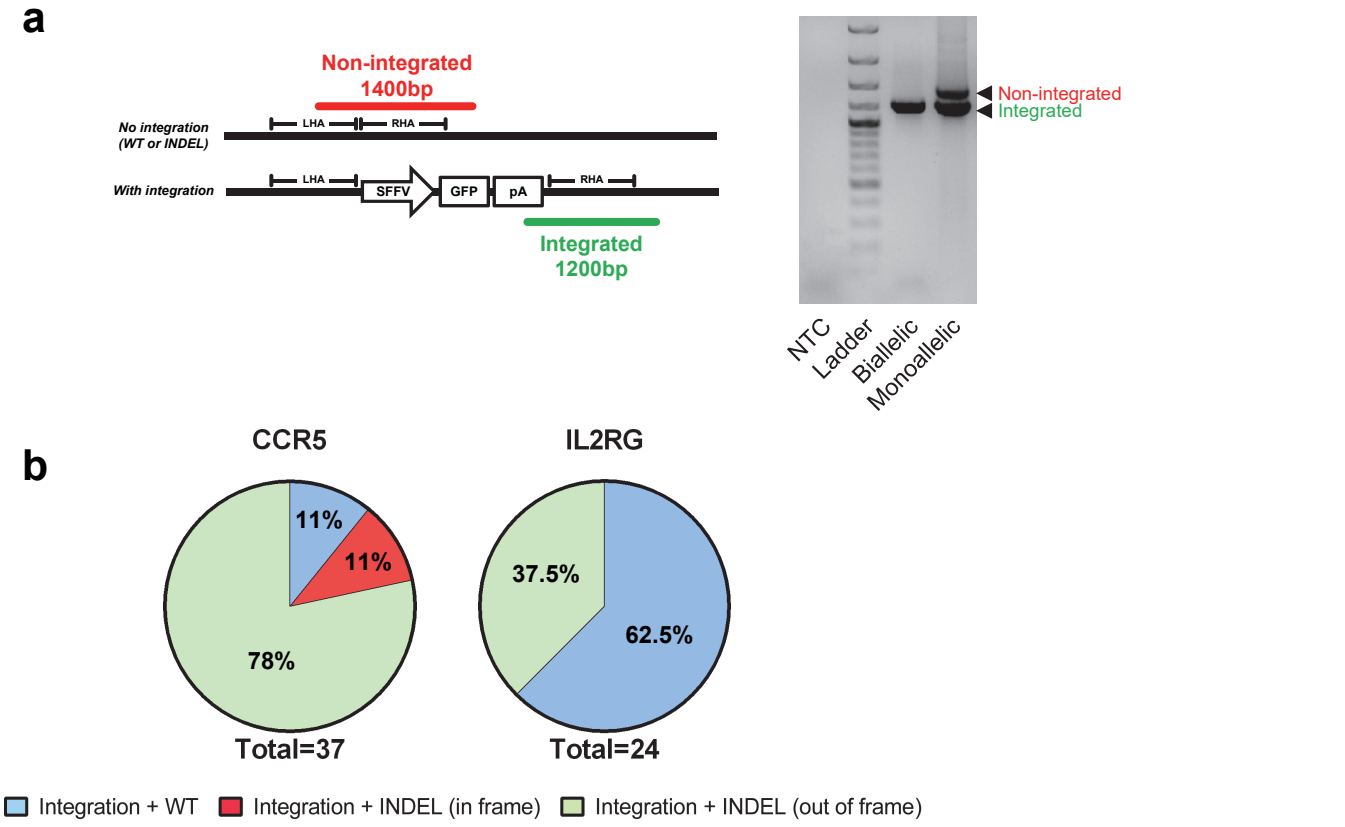


Figure 1-figure supplement 2

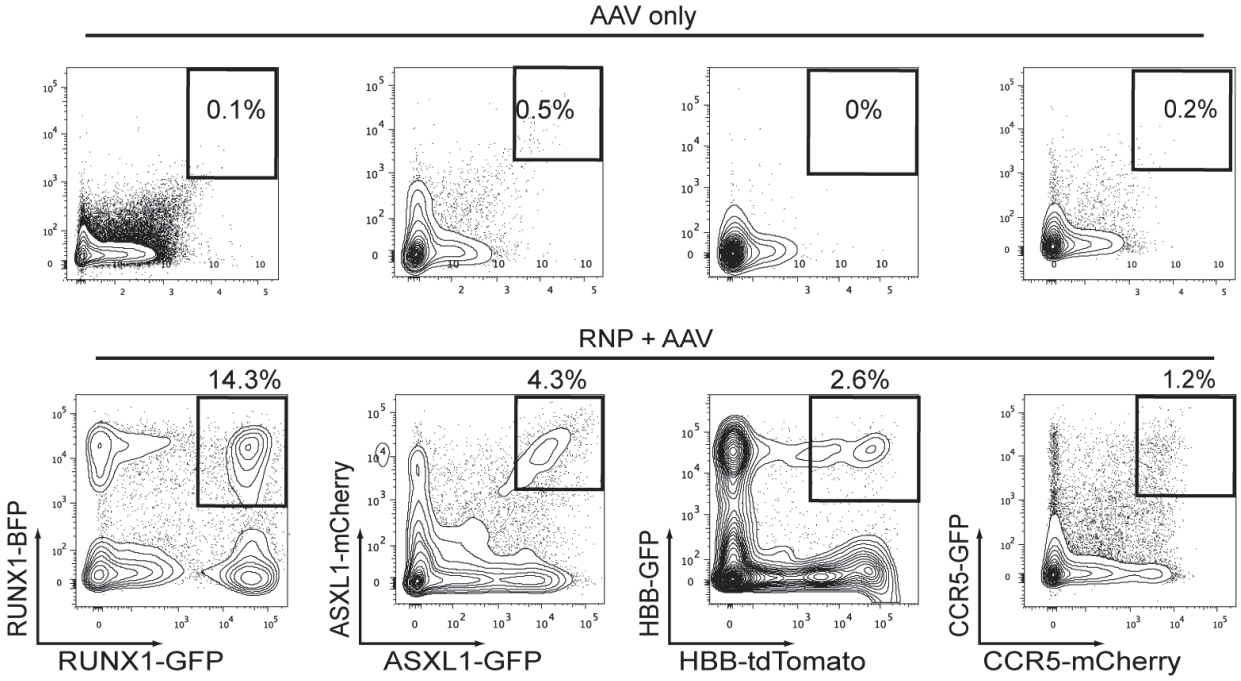


Figure 2-figure supplement 1

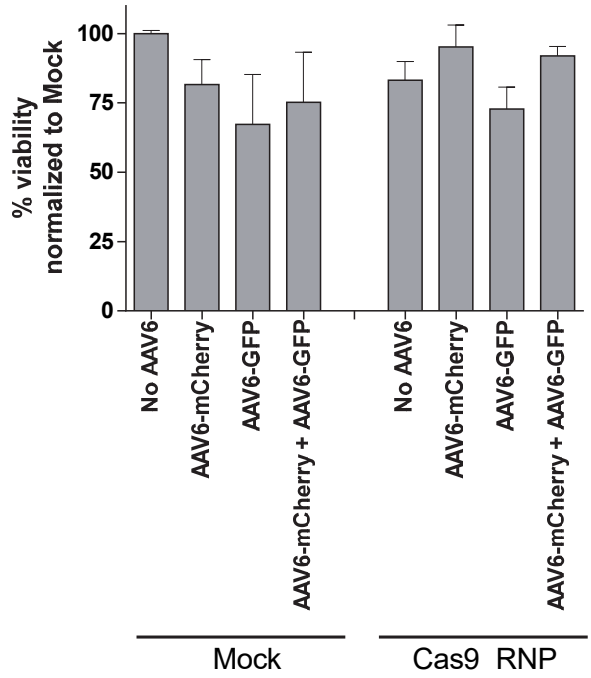


Figure 2-figure supplement 2

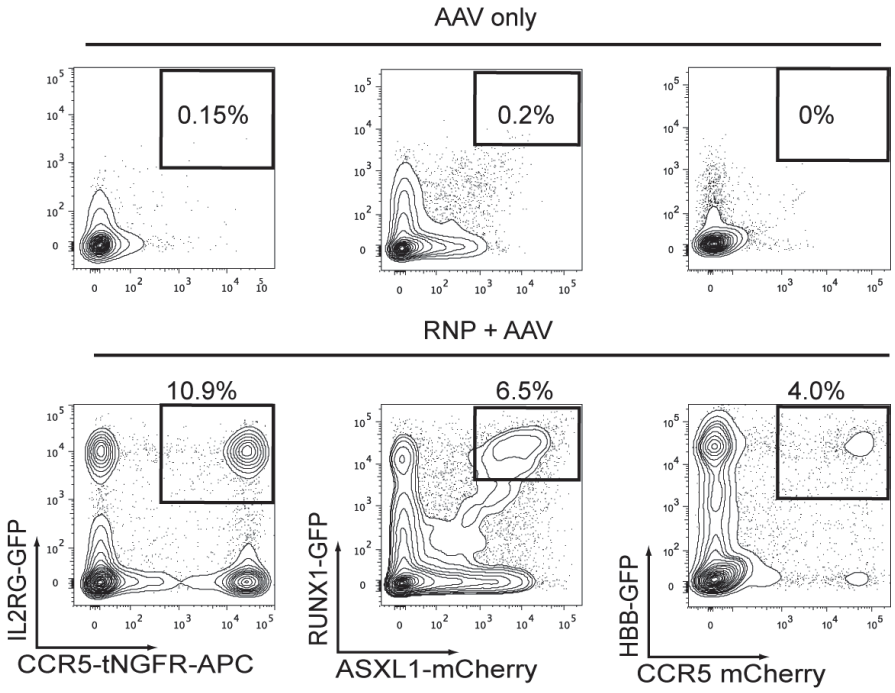


Figure 3-figure supplement 1.

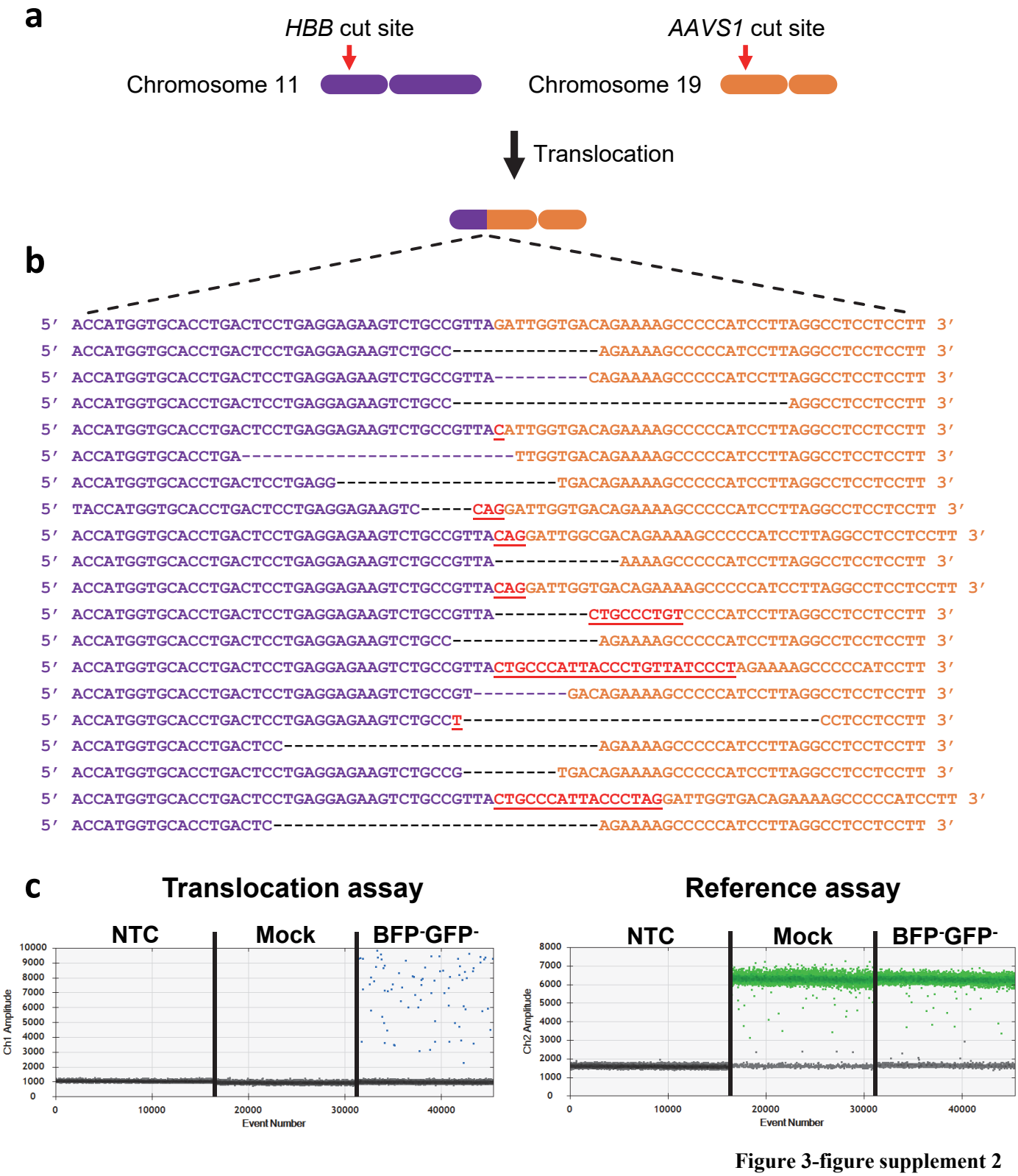


Figure 3-figure supplement 2

a

Population transplanted	Mouse #	# of transplanted cells	% Human engraftment	% B cells	% Myeloid cells	% BFP$^-$ GFP$^+$	% BFP$^+$ GFP$^-$	% BFP$^+$ GFP$^+$
Mock	1	3,0E+05	8,45	12,3	83,7	0,00	0,00	0,00
	2	3,0E+05	55,9	82,2	13,4	0,00	0,00	0,00
	3	1,2E+05	2,46	90,3	4,32	0,11	0,11	0,00
BFP$^-$ GFP$^-$	1	3,0E+05	52,7	61,4	33,8	0,00	0,03	0,00
	2	1,2E+05	44,0	81,5	11,7	0,08	0,01	0,00
BFP$^-$ GFP$^+$	1	1,2E+05	51,0	78,9	14,0	99,9	0,00	0,02
BFP$^+$ GFP$^-$	1	3,0E+05	0,31	56,7	29,0	0,00	85,2	0,98
	2	3,0E+05	6,40	66,6	28,7	0,00	96,9	0,03
BFP$^+$ GFP$^+$	1	3,0E+05	19,4	9,54	83,0	0,15	0,00	99,1
	2	3,0E+05	0,13	0,00	85,1	2,97	0,00	80,2
	3	3,0E+05	0,31	50,4	35,6	0,33	0,00	29,6

Figure 3-figure supplement 3

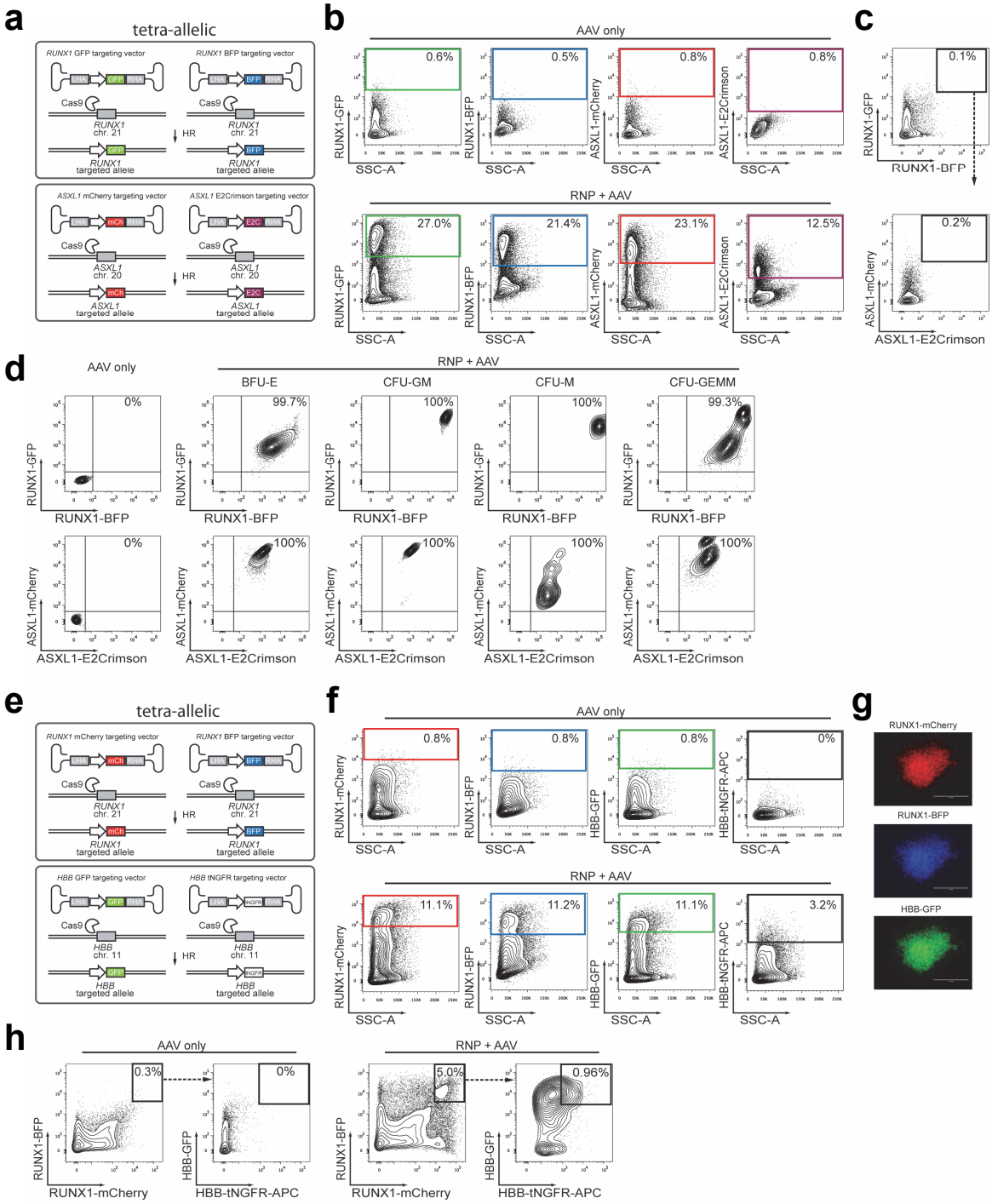


Figure 4-figure supplement 1

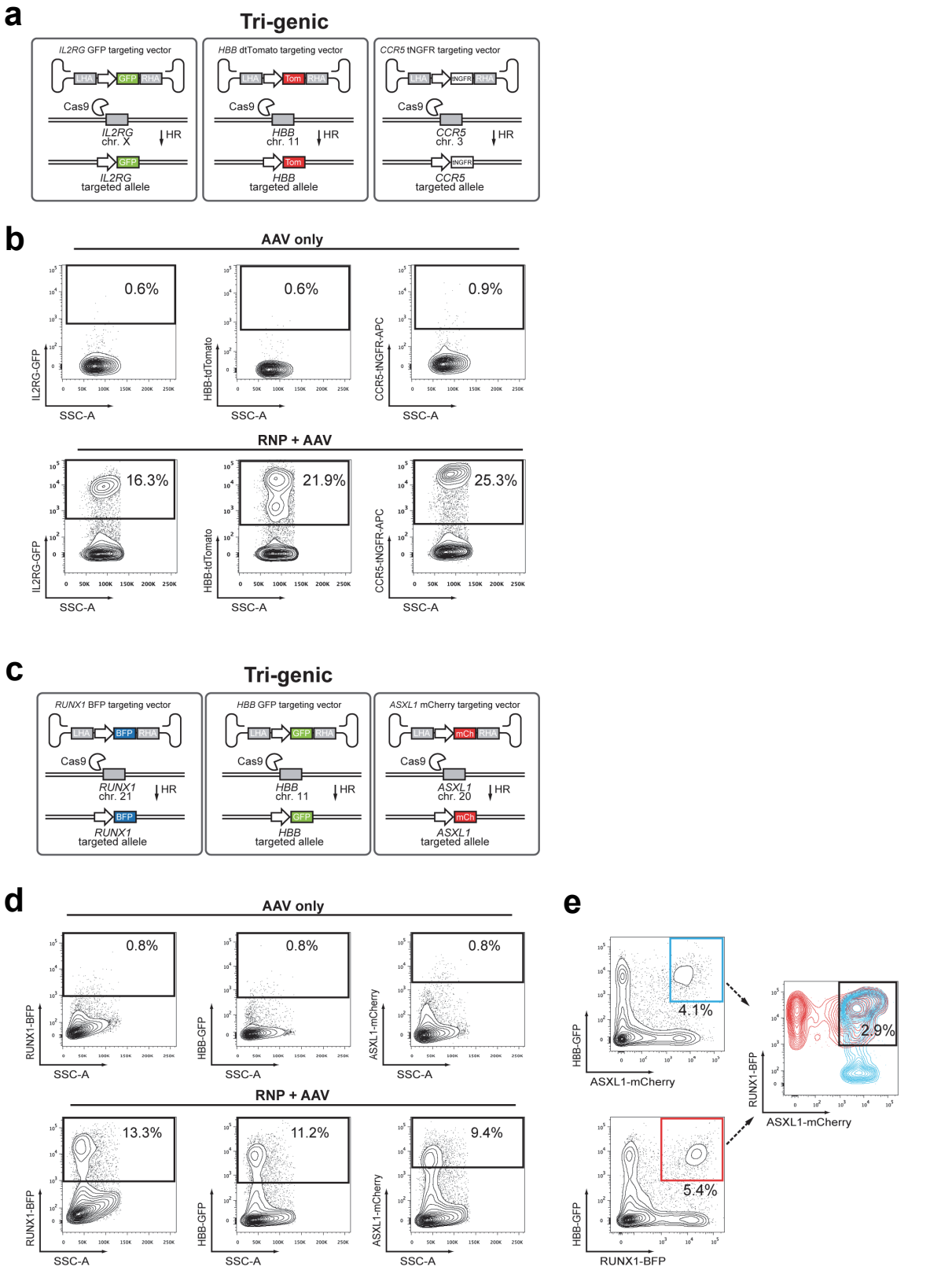


Figure 4-figure supplement 2

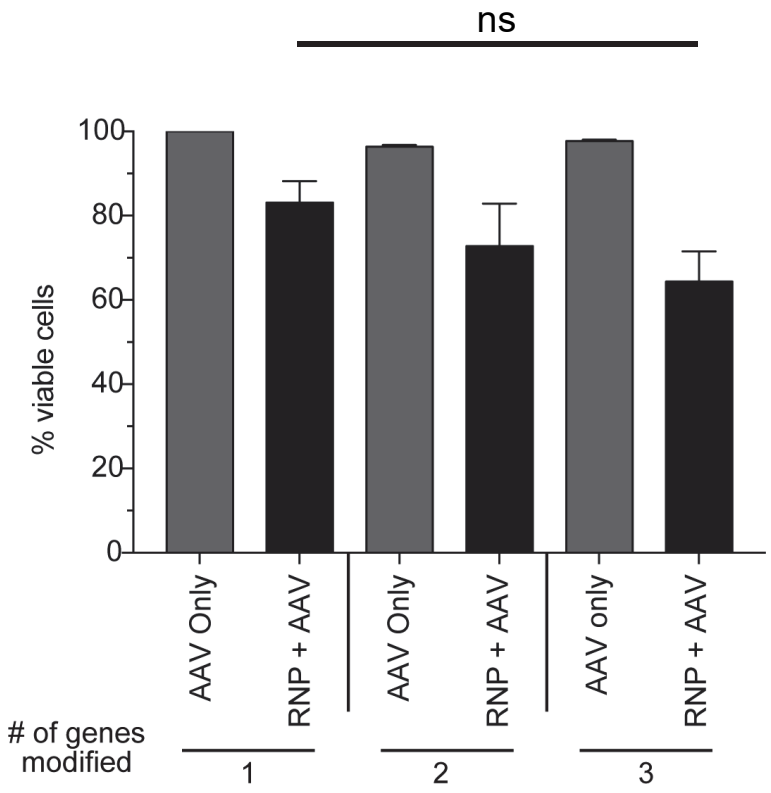


Figure 4-figure supplement 3

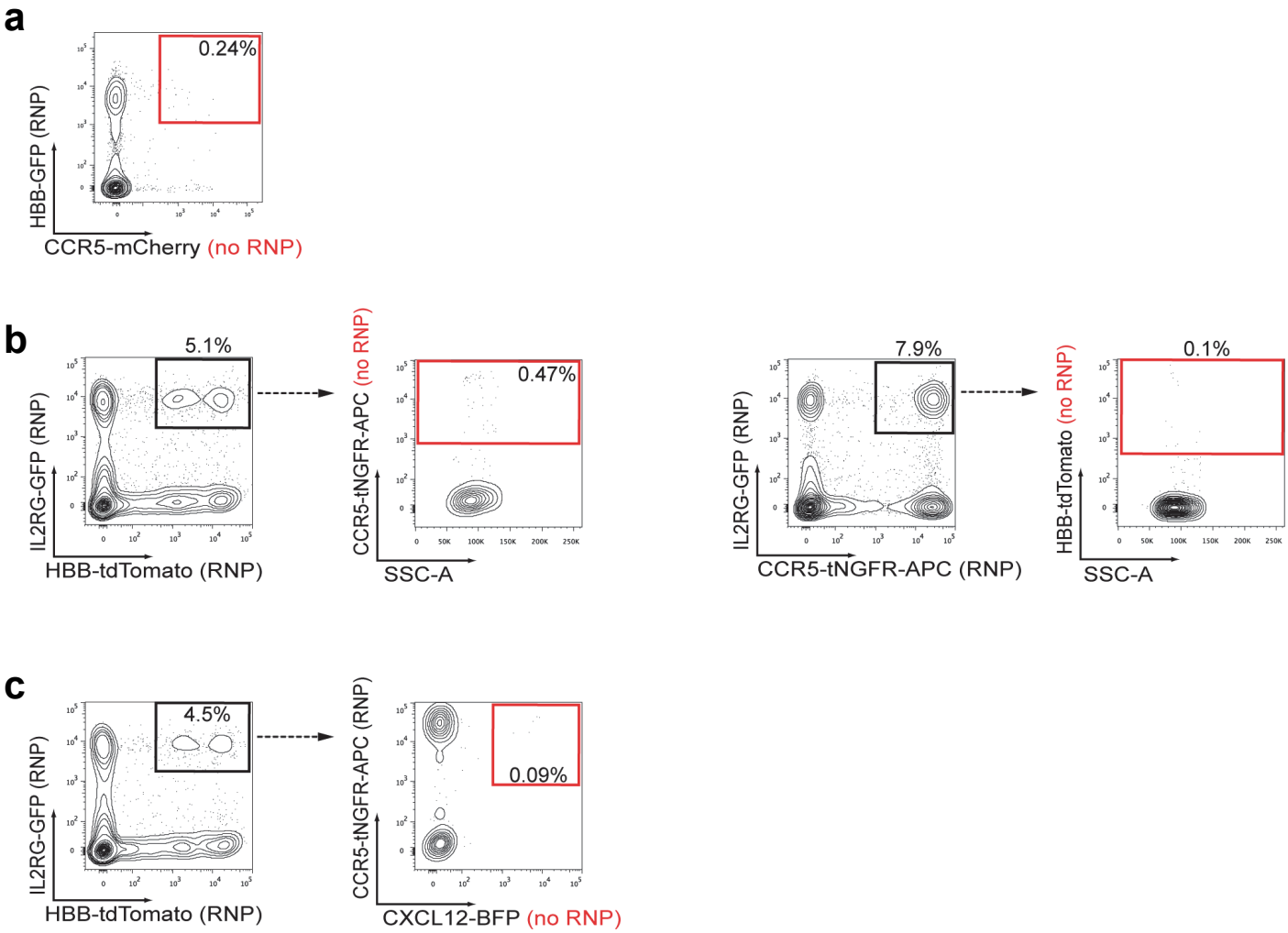


Figure 4-figure supplement 4

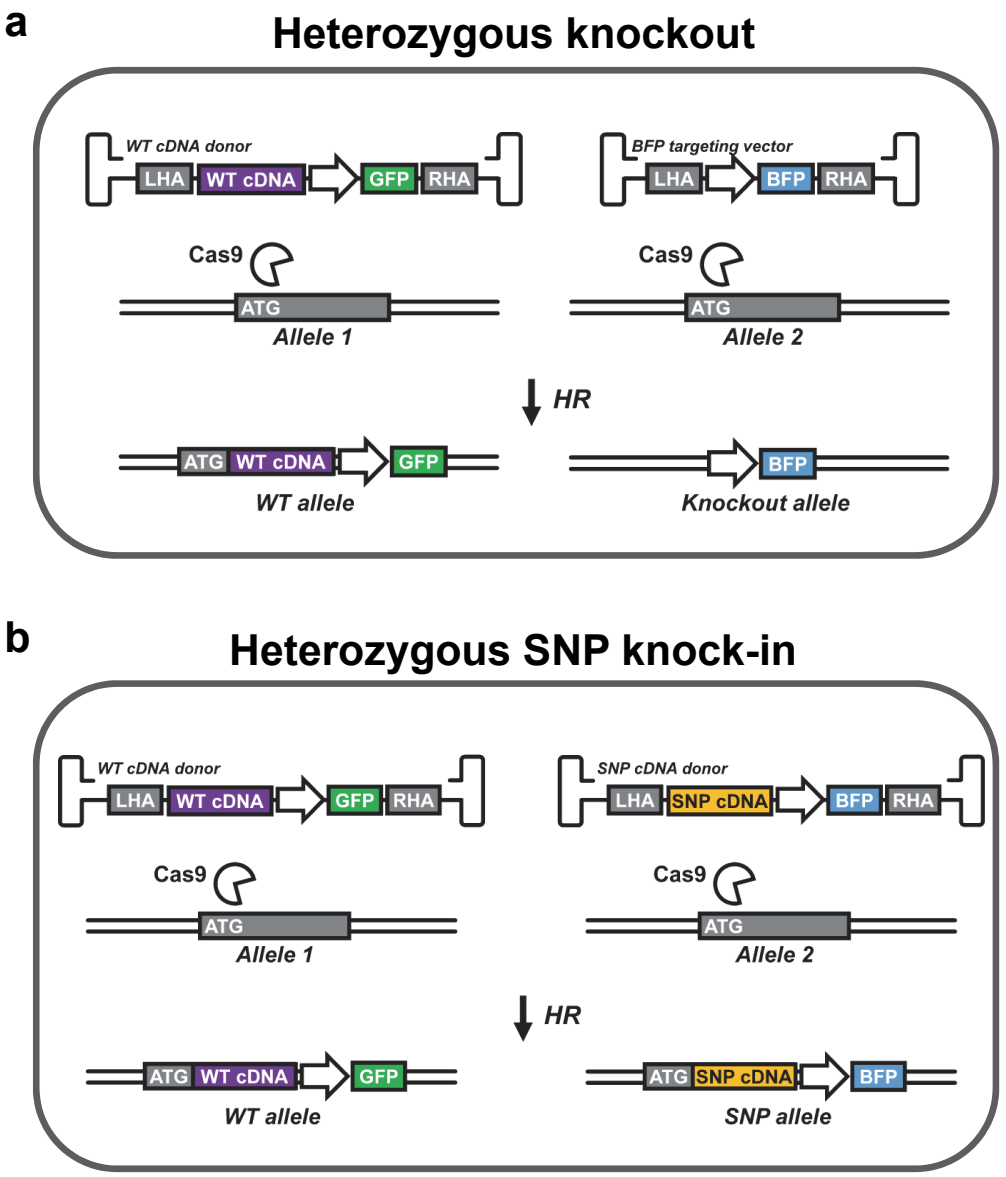


Figure 4-figure supplement 5