Stem cell-derived clade F AAVs mediate high-efficiency homologous recombination-based genome editing

Laura J. Smith[a], Jason Wright[b], Gabriella Clark[a], Taihra Ul-Hasan[a], Xiangyang Jin[a], Abigail Fong[a], Manasa Chandra[a], Thia St Martin[b], Hillard Rubin[b], David Knowlton[b], Jeff L. Ellsworth[b], Yuman Fong[a], Kamehameha K. Wong Jr.[c], and Saswati Chatterjee[a,1]

[a]Department of Surgery, Beckman Research Institute, City of Hope Medical Center, Duarte, CA 91010; [b]Homology Medicines, Inc., Bedford, MA 01730; and [c]Department of Hematology and Stem Cell Transplantation, City of Hope Medical Center, Duarte, CA 91010

Edited by Kenneth I. Berns, University of Florida College of Medicine, Gainesville, FL, and approved June 14, 2018 (received for review February 7, 2018)

The precise correction of genetic mutations at the nucleotide level is an attractive permanent therapeutic strategy for human disease. However, despite significant progress, challenges to efficient and accurate genome editing persist. Here, we report a genome editing platform based upon a class of hematopoietic stem cell (HSC)-derived clade F adeno-associated virus (AAV), which does not require prior nuclease-mediated DNA breaks and functions exclusively through BRCA2-dependent homologous recombination. Genome editing is guided by complementary homology arms and is highly accurate and seamless, with no evidence of on-target mutations, including insertion/deletions or inclusion of AAV inverted terminal repeats. Efficient genome editing was demonstrated at different loci within the human genome, including a safe harbor locus, AAVS1, and the therapeutically relevant IL2RG gene, and at the murine Rosa26 locus. HSC-derived AAV vector (AAVHSC)-mediated genome editing was robust in primary human cells, including CD34$^+$ cells, adult liver, hepatic endothelial cells, and myocytes. Importantly, high-efficiency gene editing was achieved in vivo upon a single i.v. injection of AAVHSC editing vectors in mice. Thus, clade F AAV-mediated genome editing represents a promising, highly efficient, precise, single-component approach that enables the development of therapeutic in vivo genome editing for the treatment of a multitude of human gene-based diseases.

adeno-associated virus | genome editing | homologous recombination | hematopoietic stem cells | in vivo genome editing

The ultimate goal of genetic medicine is to precisely and efficiently edit the human genome to correct somatic mutations or insert therapeutic sequences at targeted chromosomal locations. Despite significant recent progress, current genome editing technologies have yet to attain therapeutically relevant levels of efficiency and precision, particularly in vivo. The majority of current gene editing platforms utilize nuclease-mediated DNA cleavage as the first step in editing events. The efficiency and accuracy of nuclease-mediated cleavage are often variable (1), and repair of double-stranded DNA breaks by homologous recombination (HR) is low (2–11). These nuclease-induced DNA breaks are then repaired by the error-prone nonhomologous end-joining (NHEJ) pathway, which results in random nucleotide insertions/deletions (indels) at the cleavage site, potentially creating undesirable on-target mutations (12) or insertion of the entire vector genome, including the inverted terminal repeats (ITRs) (13). In contrast, DNA break repair by HR is more precise and can be engineered to accurately insert specific sequences at precise chromosomal locations, but it occurs at a lower frequency than NHEJ-mediated repair (2–11, 14–16). While nuclease-based genome editing is effective for gene disruption, the predictable correction of mutations and targeted insertion of therapeutic sequences are harder to achieve due to the combination of the low frequency of HR repair coupled with the requirement to deliver multiple components to target cells. Another significant challenge to the therapeutic translation of genome editing is that classic HR is restricted to dividing cells, thus excluding most cells in vivo, which exist in a postmitotic state (17). Nuclease-based editing platforms also carry the burden of promiscuous off-target cleavage by which double-stranded DNA breaks are created at unintended sites, resulting in the potential for genome-wide mutagenesis (18–23). Targeting of nucleases to specific genomic sites is accomplished by the use of DNA-binding domains or guide RNA, which can place constraints on the exact location of cleavage sites (23). Finally, efficient in vivo delivery of genome-editing components to target tissues remains challenging. Most platforms consist of multiple components, which are often too large to fit into a single delivery vector. To achieve genome editing in vivo, each target cell would require the delivery of multiple editing components, further reducing efficiency. Thus, while much effort has been focused upon translating genome editing to therapeutic applications, significant challenges persist. An accurate, efficient, and predictable genome editing technology based solely on HR pathways would thus represent a significant advance in the field and would allow precise targeted insertion of therapeutic sequences and gene correction without the risk of on-target mutagenesis or promiscuous off-target effects. If such an editing platform functioned efficiently in postmitotic cells in vivo and consisted of a single component

Significance

Precise genomic correction of pathogenic mutations is an attractive therapeutic strategy. Here, we show that a unique class of stem cell-derived nonpathogenic virus, hematopoietic stem cell-derived adeno-associated virus vector (AAVHSC), mediates precise genome editing at unprecedented efficiencies in primary human cells, including postmitotic cells, and in vivo. Unlike the majority of current editing platforms, genome editing by AAVHSC is uniquely based on homologous recombination (HR), requiring no exogenous nucleases. AAVHSC-mediated editing is seamless, with no evidence of on-target insertion/deletion mutations common to nuclease-based platforms. Efficient in vivo editing was achieved by i.v. injection of AAVHSC editing vectors. The combination of efficient and precise HR-based genome editing coupled with the superior in vivo transduction properties of AAV facilitates progress toward in vivo therapeutic gene editing.

Author contributions: J.W., Y.F., K.K.W., and S.C. designed research; L.J.S., J.W., G.C., T.U.-H., X.J., A.F., M.C., T.S.M., H.R., D.K., K.K.W., and S.C. performed research; L.J.S., J.W., X.J., J.L.E., K.K.W., and S.C. analyzed data; and L.J.S., J.W., and S.C. wrote the paper.

Conflict of interest statement: S.C. and L.J.S. are cofounders of Homology Medicines, Inc., and S.C., L.J.S., J.W., T.S.M., H.R., J.L.E., and D.K. hold equity in Homology Medicines, Inc.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: schatterjee@coh.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1802343115/-/DCSupplemental.

amenable to in vivo use, it would further resolve the significant challenges currently facing therapeutic genome editing.

Adeno-associated virus (AAV)–based vectors have previously been shown to mediate genome editing without the requirement for exogenous nucleases (24–29). However, the frequencies of genome editing reported were very low, generally well below 0.1%. In this study, we interrogated a class of CD34$^+$ hematopoietic stem cell-derived AAV vectors (AAVHSCs) (30), for their genome editing capacity. The AAVHSCs represent a family of natural, human, nonpathogenic, single-stranded, replication-defective AAVs that infect both dividing and nondividing cells and display broad systemic tropism in vivo (30–36). Here, we show that AAVHSCs mediate precise, efficient on-target genome editing in primary human cells and at multiple genomic locations, including therapeutically relevant genes in vitro and in vivo, without the requirement for prior nucleolytic DNA cleavage. Importantly, AAVHSC-based genome editing occurs exclusively via the HR pathway and is highly efficient in primary human cells. We further show that the precision of on-target editing is highly accurate, with no evidence of genomic scarring, indels, or incorporation of a residual viral footprint. We demonstrate a highly precise, exclusively HR-based, nuclease-free genome editing platform that functions at unparalleled efficiencies in primary cells and in vivo and is composed of a single component with built-in in vivo delivery properties.

Results

AAVHSCs Mediate Efficient Genome Editing in Human Cells in the Absence of Exogenous Nucleases. We designed a reporter-based editing assay to measure the targeted genomic insertion of a promoterless green fluorescent protein (GFP) into the safe harbor locus, AAVS1, located on human chromosome 19 (37), such that expression would be dependent upon accurate insertion downstream of a chromosomal promoter (Fig. 1*A*). The editing vector genome was bounded at both ends by AAV ITRs, consisted of a promoterless GFP ORF immediately downstream of a splice acceptor (SA) and T2A (SA/T2A) sequence followed by a polyadenylation (pA) signal, and was flanked bilaterally by 800-bp homology arms (HAs) complementary to the PPP1R12C gene sequence on human chromosome 19 within the safe harbor locus AAVS1 (37). This vector, termed PPP1R12C-GFP, was designed to insert the GFP ORF into intron 1 of the PPP1R12C gene (Fig. 1*A* and *SI Appendix*, Table S1). We reasoned that accurate editing of the promoterless GFP cassette into the intron downstream of the PPP1R12C promoter would result in GFP expression driven by the endogenous PPP1R12C promoter (Fig. 1*A*).

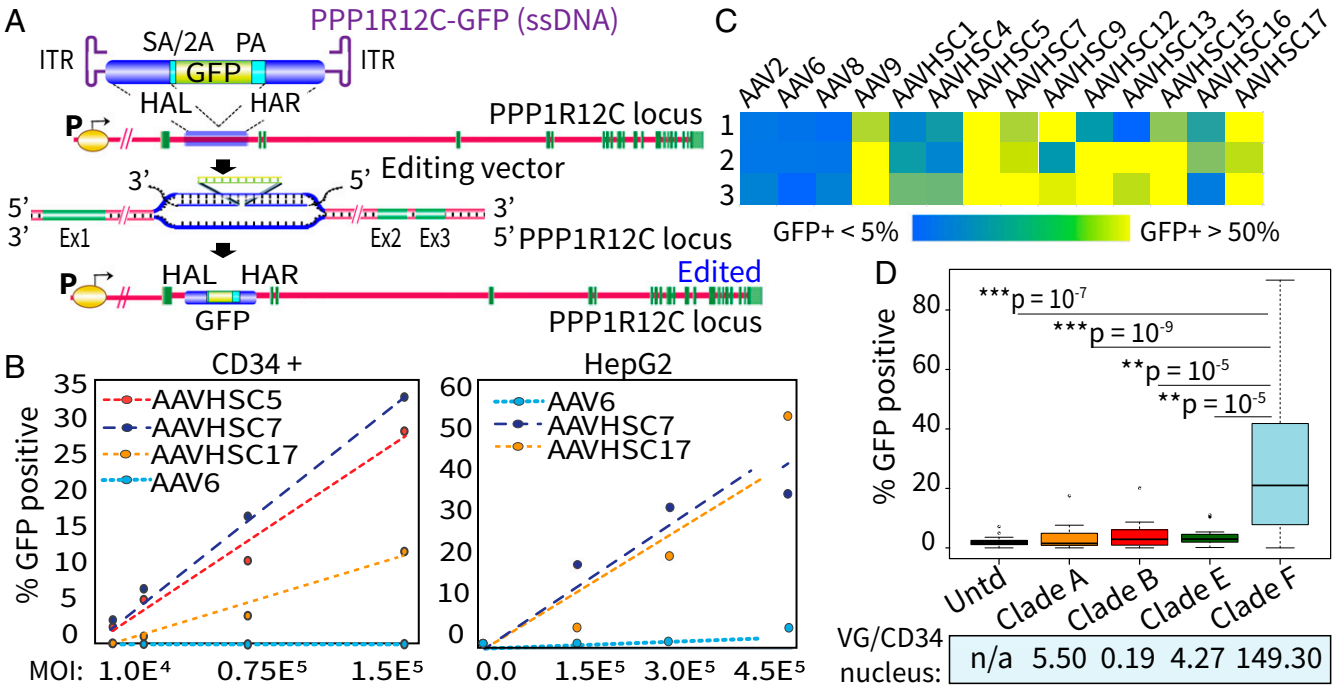

Fig. 1. Genome editing of a promoterless GFP into intron 1 of the human PPP1R12C gene. (*A*) Schema for genome editing assay. (*Upper*) Map of the PPP1R12C-GFP editing vector genome. In line 1, the insert cassette consists of a GFP ORF (green) preceded by SA/T2A (SA/2A) sequences and followed by a pA (PA) signal (cyan), and it is flanked on either side by 800-bp HAs (blue) complementary to chromosomal sequences in PPP1R12C. HAL, left HA; HAR, right HA. The entire editing construct is bounded by AAV2 ITRs (purple). In line 2, the PPP1R12C gene is depicted. Promoter (P) is shown in yellow, exons (Ex) are shown in dark green, and introns are shown in pink. Line 3 depicts a hypothetical HR schema. Line 4 depicts the edited PPP1R12C gene showing the insertion of the GFP cassette into intron 1. (*B*) Dose–response curves correlating genome editing with the MOI of the PPP1R12C-GFP editing vector. The editing vector genome was packaged in the capsids noted within each box. (*Left*) Editing of primary human CD34$^+$ cells. (*Right*) Editing of the human hepatocellular carcinoma cell line HepG2. (*C*) Heat map showing genome editing efficiency of targeted insertion of GFP into PPP1R12C by AAV serotype (columns) in different human cell types (rows) as follows: (1) transformed cell lines, (2) primary human cells, and (3) immortalized LCLs carrying genetic mutations. Values represent specific GFP expression as assessed by flow cytometry. GFP expression was assessed only in viable cells, as determined by DAPI exclusion. Data shown are aggregates of >906 individual measurements. Cell type details are shown in *SI Appendix*, Fig. S1. (*D*) Comparison of median genome editing efficiency across AAV clade A (AAV6), clade B (AAV2), clade E (AAV8), and clade F (AAV9) (AAVHSCs). Editing efficiencies were measured by GFP expression by flow cytometry. Cells were transduced at a MOI of 1.5E5 and evaluated by flow cytometry at 48 h. Sample sizes for the experimental groups are as follows: untreated (Untd), $n = 25$; AAV6, $n = 33$; AAV2, $n = 17$; AAV8, $n = 19$; AAV9, $n = 30$; and AAVHSCs combined, $n = 390$. AAVHSC represents data compiled from AAVHSC1, AAVHSC4, AAVHSC5, AAVHSC7, AAVHSC9, AAVHSV12, AAVHSC13, AAVHSV15, AAVHSV16, and AAVHSC17. Outliers are represented by individual circles. Significance was determined by a paired two-tailed *t* test using AAVHSCs as the comparison reference. The vector genomes (VG) were quantitated in nuclei purified from AAV-treated CD34$^+$ cells 48 h posttreatment. The number of VG per nucleus was determined by real-time PCR for GFP and the housekeeping gene hApoB. Values shown are averages of three replicates per transduction and three transductions with each AAV vector.

The promoterless GFP editing vector was packaged in AAVHSC5, AAVHSC7, AAVHSC17, and AAV6 capsids. A titration of the multiplicities of infection (MOIs) revealed a linear relationship between GFP expression and vector concentration for each AAV serotype tested in both primary CD34$^+$ cells and the hepatocellular carcinoma cell line HepG2 (Fig. 1*B*), indicating that GFP expression correlated directly with the cell/vector ratio in a dose-dependent manner. The percentage of cells expressing GFP and the slopes of the titration curves revealed that the AAVHSCs mediated targeted insertion of GFP more efficiently than AAV6 in both cell types (Fig. 1*B*). All subsequent in vitro experiments described here were performed at a MOI of 1.5E5.

We next compared levels of GFP expression after insertion into intron 1 of PPP1R12C following transduction by AAVHSCs, AAV2, AAV6, AAV8, and AAV9 vectors, in over 906 individual measurements across an array of human cells (Fig. 1*C* and *SI Appendix*, Fig. S1), including primary human CD34$^+$ HSCs, liver sections, hepatic sinusoidal endothelial cells, myoblasts, and EBV-immortalized human B lymphoblastoid cell lines (LCLs) (*SI Appendix*, Fig. S1). Cumulative GFP expression was higher in cells treated with clade F HSC-derived AAVs, including AAVHSCs and AAV9, up to >50% across a spectrum of human cell types (Fig. 1*C* and *SI Appendix*, Fig. S1). In contrast, cells treated with AAV2 (clade B), AAV6 (clade A), and AAV8 (clade E) displayed lower GFP expression (Fig. 1*C* and *SI Appendix*, Fig. S1), in concordance with previous reports (24–29). Importantly, evaluation of cell viability by staining with 4′,6-diamidino-2-phenylindole (DAPI) followed by flow cytometric analysis showed no difference between AAVHSC-treated cells and controls, indicating that genome editing with AAV vectors was not toxic (*SI Appendix*, Table S2). This was further confirmed by morphological analysis.

A comparison of the median cumulative GFP expression revealed that clade F AAVs edited at a significantly higher efficiency, with a median efficiency of 24% across all human cell types studied, compared with clade B (AAV2, 2.12%), clade A (AAV6, 0.43%), and clade E (AAV8, 1.7%) (Fig. 1*D* and *SI Appendix*, Fig. S2*A*). While the AAVHSCs edited at a higher efficiency than AAV2, AAV6, and AAV8, we observed variability in GFP expression between different AAVHSCs (*SI Appendix*, Fig. S2*A*). Interestingly, AAVHSCs and AAV9 were found in human CD34$^+$ cells, share significant sequence homology, map to clade F (30), and displayed similar enhanced genome editing properties. Thus, we conclude that stem cell-derived clade F AAVs represent a powerful class of vectors that mediate efficient genome editing and targeted insertion at specified chromosomal locations, at unparalleled efficiencies in primary human cells in the absence of exogenous nucleases.

Both genome editing and gene transfer by AAV vectors must undergo common early cell interaction processes, including receptor-mediated entry, translocation to the nucleus, nuclear entry, and uncoating before vector genomes are released and available for either transgene expression, in the case of gene transfer, or genome editing. Thus, we compared the efficiency of genome editing, as measured by promoterless GFP expression, with that of gene transfer for the different AAV serotypes (*SI Appendix*, Fig. S2*B*). Results revealed a direct correlation between transduction and editing (R^2 = 0.8337), suggesting that specific virus–cell interactions may represent rate-limiting steps that are common to both processes and indicating that the overall transduction efficiency, as measured by transgene expression, may be a contributing factor in editing efficiency in vitro. Since serotype-dependent differences have been shown to play a role in nuclear entry (38), we quantitated intranuclear vector genomes in purified nuclei following transduction of CD34$^+$ cells with different AAV clades (Fig. 1*D*). Results revealed that the number of intranuclear vector genomes was higher for clade F AAVs [average = 149.29 vector genomes (vg) per nucleus] at 48 h posttransduction, compared with AAV vectors belonging to clade A (5.49 vg per nucleus), clade B (0.18 vg per nucleus), and clade E (4.26 vg per nucleus). Thus, clade F AAVs are over 27-fold more efficient at delivering genomes to the nucleus compared with AAV belonging to clades A, B, and E. Additional factors, including second-strand synthesis and genome concatamerization, may also play roles in guiding editing efficiencies but remain as yet undefined.

Molecular Characterization of AAVHSC-Mediated Genome Editing in CD34$^+$ Cells. We next evaluated the on-target precision and efficiency of HR at the allelic level in primary CD34$^+$ cells treated with the promoterless PPP1R12C-GFP editing vectors designed to insert the GFP ORF into intron 1 of the PPP1R12C gene in AAVS1. Flow cytometric analysis of AAVHSC-treated CD34$^+$ cells revealed distinct GFP-expressing populations. In a representative experiment, CD34$^+$ cells treated with AAVHSC5, AAVHSC7, AAVHSC12, and AAVHSC17 expressed GFP in 47.4%, 41.8%, 58.1%, and 48.9% of cells within 48 h, respectively (Fig. 2*A*). In contrast, no specific GFP expression was observed following treatment with the promoterless AAV6 PPP1R12C-GFP editing vector (Fig. 2*A*). Similarly, AAV2 and AAV8 showed low levels of GFP expression in primary CD34$^+$ cells (*SI Appendix*, Fig. S1). Donor variability in transduction efficiencies of primary cells, including CD34$^+$ cells, has previously been reported (32, 39) and may account for differences observed between experiments with cells from different donors.

To confirm the accuracy of on-target editing, we developed a targeted integration (TI) assay (Fig. 2*B* and *SI Appendix*, Fig. S3). A GFP-specific primer and a chromosome-specific primer complementary to the genomic sequences outside the 5′ HA were used to amplify a 1.7-kb fragment, representing accurate insertion of the GFP ORF into intron 1 of PPP1R12C (Fig. 2*B* and *SI Appendix*, Fig. S3*A*). The correctly sized 1.7-kb band was amplified only from AAVHSC-treated, but not untreated, HepG2 and CD34$^+$ cells (Fig. 2*B*). Densitometric analyses of the gels revealed an average 20-fold increase of band intensity of the TI-specific amplicon in lanes containing amplified DNA from cells treated with the editing vector compared with untransduced and no-template controls (*SI Appendix*, Fig. S4*A*), and an eightfold increase over untransduced controls from CD34$^+$ day 39 gel (Fig. 2*B* and *SI Appendix*, Fig. S4*B*). Specific TI amplicons were observed within 24 h of treatment and were readily detected as late as 39 d posttreatment in CD34$^+$ cells maintained in culture in the absence of selective pressure, suggesting that the genome editing event occurred rapidly and was stable over time (Fig. 2*B*). GFP expression was assayed in these long-term cultures in parallel on days 20 and 39 (*SI Appendix*, Table S3). Notably, due to the asymmetrical differentiation of primary CD34$^+$ cells in culture into different hematopoietic lineages with varying lifespans, long-term cultures do not accurately reflect the initial cultures. TI assays with primers spanning the 3′ edited chromosomal junction region also revealed amplification of the correctly sized bands (*SI Appendix*, Fig. S3*A*). Importantly, control amplifications of untransduced cell DNA spiked with an excess of purified PPP1R12C-GFP editing vector genomes did not show any amplicons (*SI Appendix*, Fig. S3*B*, *Lower Left*, lane 5), excluding the possibility of cross-over amplification with episomal vector DNA.

To confirm that the TI amplicons represented the correct targeted insertion product, the 1.7-kb amplicon spanning the 5′ chromosomal junction, including the left HA and into the SA/2A-GFP insert, was subjected to Sanger sequencing (*SI Appendix*, Fig. S5 *A* and *B*). Since the high GC content of the palindromic AAV ITRs is known to be difficult to sequence through, we developed a method that keeps the ITRs sufficiently denatured so as to reproducibly obtain complete sequence reads (40). Sequence analysis from edited CD34$^+$, K562, and HepG2 cells revealed that targeted insertion was precise, with the chromosomal sequences being contiguous with HA sequences, followed by the SA/T2A and the GFP ORF. There was no evidence of sequence alterations in either the chromosome or junction regions or in the insert (*SI Appendix*, Fig. S5 *A* and *B*). To further confirm the accuracy of editing, we sequenced both junction regions, including the target chromosome-HA as well as the junction

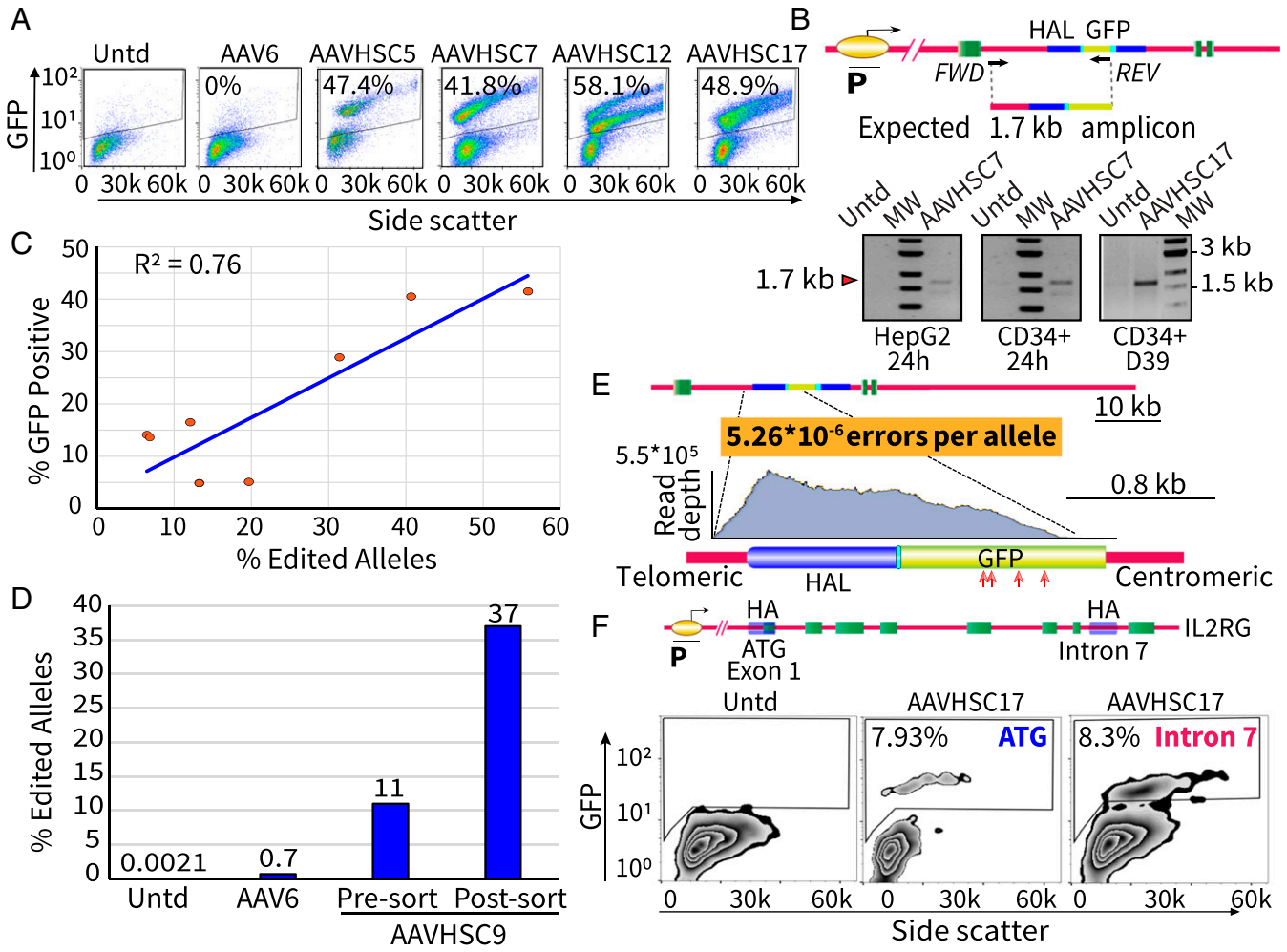

Fig. 2. AAVHSCs mediate precise and efficient genome editing in primary human cells. (*A*) Flow cytometric analysis of GFP expression in cord blood CD34$^+$ cells following editing of the promoterless GFP ORF into intron 1 of PPP1R12C. Cells were transduced at a MOI of 1.5E5 and assayed for GFP expression after 48 h. The fraction of specific GFP expression is noted in each flow profile. Untd, untreated. (*B*) Schema depicting the edited PPP1R12C locus showing targeted insertion of GFP, the expected 1.7-kb TI amplicon, and the location of primers for the TI assay. Gels show TI amplicons (red arrow) specific for GFP insertion into intron 1 of PPP1R12C in HepG2 and CD34$^+$ cells at 24 h and 39 d posttransduction in vitro. *FWD*, forward; HAL, left HA; MW, molecular weight marker; P, promoter; *REV*, reverse. (*C*) Correlation of GFP expression in edited cells with molecular ddPCR-based quantitation of edited alleles. CD34$^+$ cells were treated with AAVHSC17 PPP1R12C-GFP editing vector at a MOI of 1.5E5. Flow cytometry-based GFP expression and ddPCR-based allele quantitation were measured from the same samples after 48 h. (*D*) Enrichment of edited alleles in flow-sorted K562 cells. K562 cells were treated with the AAVHSC9 PPP1R12C-GFP editing vector at a MOI of 150,000. Cells were assessed for the frequency of edited alleles before and 3 wk after enrichment of GFP-expressing cells. (*E*) NGS analysis of the edited PPP1R12C genomic region from edited CD34$^+$ cells. Shown is the NGS read depth (*y* axis) and the covered chromosomal region (*x* axis), including HAs and the GFP insert. The fidelity of editing and errors per allele is noted, showing seamless editing with no inserted viral sequences being detected. The locations of errors are denoted by red arrows under the map. Each arrow signifies a single error. (*F*) Targeted editing of the promoterless GFP cassette into two locations within the human IL2RG gene in primary human CD34$^+$ cells. The map depicts the IL2RG gene. Insertion locations are at either the ATG or in intron 7 (HA; depicted in blue). Flow analyses of GFP expression following insertion of the GFP ORF at either the initiation codon or in intron 7 are shown. Expression is driven by the chromosomal IL2RG promoter. Percent-specific GFP is noted within each flow profile.

between each HA and GFP insert using chromosome-specific primers complementary to regions external to HAs (*SI Appendix*, Fig. S5*C*). In concordance with the previous results, editing was found to be seamless at all junctions with no indel mutations or sequence disruptions. Importantly, no elements of the viral genomic backbone, including the ITRs, were found inserted into the genome, suggesting that HR could be involved. These results provide sequence confirmation of successful and precise editing of the promoterless GFP cassette into the intended chromosomal location. Importantly, this was observed in primary human cells and cell lines, including human CD34$^+$ cells, which predominantly existed in a resting state at the time of treatment (36).

To quantitate the fraction of alleles that were successfully edited, we developed a droplet digital PCR (ddPCR)-based allele quantitation assay (Fig. 2*C* and *SI Appendix*, Fig. S6). In this assay, cellular genomes are diluted to ∼30 cells per 20-μL reaction, a concentration at which both the cellular and vector genomes are within the linear range of detection in these experiments. To detect editing, we employed multiplexed Taqman assays specific for GFP and the chromosomal location, respectively. A duplex Taqman assay was performed with each genomic DNA sample within an oil emulsion that partitions each reaction such that each droplet is either positive or negative for GFP and the genome. Each droplet is read out by distinguishing PCR-positive droplets from PCR-negative droplets within each channel. Free vector partitions to the upper left quadrant (*SI Appendix*, Fig. S6*A*, blue dots), and unedited chromosomes partition to the lower right quadrant (*SI Appendix*, Fig. S6*A*, green dots). Edited genomes containing both the GFP-specific signal and the chromosome-specific signal *in cis* on a single

molecule of DNA are represented in the partition to the upper right quadrant (*SI Appendix*, Fig. S6*A*, orange dots), as well as the possible coincidence of genome and free vector. To control for the coincidence of GFP and genome probes, the number of expected double-positive droplets generated by chance was subtracted from the observed number of double-positive droplets. A standard curve generated for this assay (*SI Appendix*, Fig. S5*B*) shows a correlation coefficient of 0.9717. *SI Appendix*, Table S4 provides the complete statistical analysis of the allele quantitation assay. To confirm the quantitation of editing and to control for error due to sampling, each DNA sample was also analyzed following restriction digestion with an enzyme that cleaves between the GFP and chromosome-specific probe-binding sites. This artificial separation of the editing signal from the genome signal results in the independent partitioning of each probe target (*SI Appendix*, Fig. S7*A*). *SI Appendix*, Fig. S7*B* and Table S5 show that following restriction digestion, the signal in the upper right quadrant is fully resolved into the free vector and free locus signals, indicating that the edited allele signal (*SI Appendix*, Fig. S7*B*, orange dots) accurately represented the binding of both vector and locus probes to the same piece of DNA (*SI Appendix*, Fig. S7*B* and Table S5). Thus, we conclude that this ddPCR-based allele quantitation assay accurately measures edited chromosomes.

To determine if GFP expression in edited cells correlated with the frequency of edited alleles detected by ddPCR, we treated $CD34^+$ cells with the AAVHSC17 PPP1R12C-GFP editing vector. Results revealed that GFP expression, as measured by flow cytometry, was highly correlated with edited alleles ($R^2 = 0.76$) (Fig. 2*C*). A detailed statistical analysis is shown in *SI Appendix*, Table S6. These results validated that GFP detection by flow cytometry was an accurate measure of genome editing. To further confirm this, we tested whether flow-sorted, GFP-positive, AAVHSC-treated cells would display an enrichment of edited alleles. K562 erythroleukemia cells were treated with the AAVHSC9 PPP1R12C-GFP editing vector. Allele quantitation by ddPCR before and after GFP-based flow sorting revealed an enrichment of edited alleles, with 37% of alleles being edited 6 wk after transduction compared with 11% before sorting, indicating the stability of the AAVHSC-mediated genome editing over time (Fig. 2*D*). Given that K562 represents an aneuploid cell line, and the possibility of monoallelic editing, these results confirmed that GFP expression correlated well with the frequency of edited alleles and was stable over time in culture without selective pressure.

Editing Is Seamless with No Evidence of Genomic Scarring. Although HR can be achieved at a low range of efficiencies with nuclease-based genome editing platforms in the presence of donor molecules (16, 41), the vast majority of edited alleles undergo repair via the NHEJ pathway, resulting in a spectrum of on-target genomic indel mutations (42). Thus, the presence of genomic scars such as indels or insertion of AAV ITRs in edited genomes would then be indicative of on-target mutations and utilization of non–homology-directed repair pathways. To evaluate the presence of on-target genomic alterations across millions of alleles, we performed next-generation sequencing (NGS) of TI amplicons spanning >2 kb of the PPP1R12C target site in human $CD34^+$ cells edited with the PPP1R12C-GFP editing vector. To measure indel frequency and eliminate primer-dimer reads, we first identified our primers in each read. We then checked that each read contained at least 50 bases beyond the 3′ end of each of the genomic primers to ensure proper alignment to the reference genome using Bowtie 2 alignment software. A total of 1,901,904 demultiplexed reads satisfied this criterion. These reads were used to identify indels by comparing each sequence with those in the reference sequence, excluding natural variants in the human genome (dbSNP 147). Multiple alignment analysis for specific reads was done using the Geneious iterative *k*-mer multiple alignment tool (Geneious 6.1.7) (43). Of $>1.9E^6$ reads aligned with the genomic PPP1R12C target site, only four sequences displayed alterations or differences, compared with a reference target sequence, resulting in an on-target error rate of $<4.21E^{-6}$ indels per read (Fig. 2*E*), which is notably lower than on-target error rates generated by other editing platforms. The four indels observed were all within the GFP ORF and are denoted by red arrows (Fig. 2*E*). No recombination scars were observed at either the ends of the HAs or the insertion junctions.

Importantly, NGS analysis further revealed the complete absence of reads showing incorporation of AAV ITRs at the edited site, demonstrating the absence of insertion of viral backbone sequences into the genome. To confirm that AAV ITRs could actually be identified by NGS analysis, we spiked K562 cell DNA with a 1,000-fold excess of ITR-containing plasmids before NGS analysis. Results showed that ITRs were indeed detected in these libraries, with >34,000 and >8,000 reads covering at least 50 bp of the 5′ and 3′ ITRs, respectively (*SI Appendix*, Fig. S8). A small minority of reads spanned the entire ITR, with the majority of sequences dropping off after ∼60 bp. Thus, if ITRs were indeed incorporated at editing sites, at least 50 bp of each ITR sequence would have been detectable in the NGS analysis. NGS analysis also confirmed the findings of accurate and seamless editing as determined by Sanger sequencing (*SI Appendix*, Fig. S5). These results established that AAVHSC-mediated genome editing is highly precise and occurs without any concurrent on-target mutations commonly associated with DNA strand breaks and is consistent with the utilization of an HR editing pathway.

AAVHSCs Mediate Genome Editing at the Human IL2RG Locus. To confirm that AAVHSC-mediated genome editing was not limited to the single PPP1R12C locus, we tested editing of other loci within the human genome. We evaluated targeted insertion of the promoterless GFP ORF within two locations in the clinically relevant human IL2RG locus, mutations of which cause X-linked SCID disease. We evaluated editing at the ATG start codon and also within intron 7, such that GFP expression would be driven by the chromosomal IL2RG promoter (44).

In both editing vectors, the promoterless GFP ORF cassette was bilaterally flanked by 800-bp HAs that specified the chromosomal insertion location (*SI Appendix*, Table S1). For insertion at the start codon, no SA/T2A sequences were utilized. The editing vector targeting the intron included a SA/T2A sequence immediately upstream of the GFP ORF. Flow cytometric analysis of primary human $CD34^+$ cells treated with AAVHSC17-IL2RG revealed GFP expression in 7.93% of cells with insertion at the ATG start codon and 8.3% with insertion in intron 7 (Fig. 2*F*). Locus-dependent differences in gene editing efficiencies have been described and likely account for the differences in efficiency observed between the PPP1R12C and IL2RG loci (45). Thus, we conclude that AAVHSC-mediated editing of the IL2RG gene is potentially therapeutic. Similarly, we have also observed successful editing at two locations within the human factor VIII gene.

AAVHSCs Mediate Nucleotide Replacement. With current genome editing technologies, correction of point mutations is challenging. Instead of correcting mutations, DNA breaks repaired via the NHEJ pathway result in unpredictable on-target indels and the possible introduction of new mutations. Hence, we tested the potential of AAVHSC vectors to accurately correct point mutations at the nucleotide level. A dinucleotide editing vector was used to alter a wild-type "TA" sequence to an "AT" in intron 1 of PPP1R12C (nucleotide replacement vector; *SI Appendix*, Table S1). This dinucleotide change results in the conversion of a naturally occurring NheI restriction site in PPP1R12C intron 1 to a SphI restriction site (Fig. 3*A*). In addition to the dinucleotide replacement, the nucleotide replacement editing vector was designed to insert a 12-bp linker 245 bp downstream of the NheI site (Fig. 3*A*) to create a primer-binding site for the insert-specific primer for the TI PCR assay (Fig. 3*A*). A control vector was designed to only insert the downstream 12-bp primer-binding site, but not to alter the NheI site. Both the nucleotide replacement and control vectors had a 583-bp left HA 5′ to the NheI site and an 800-bp HA downstream of the 12-bp linker. The total AAV vector genome size was 2,011 bp (*SI Appendix*,

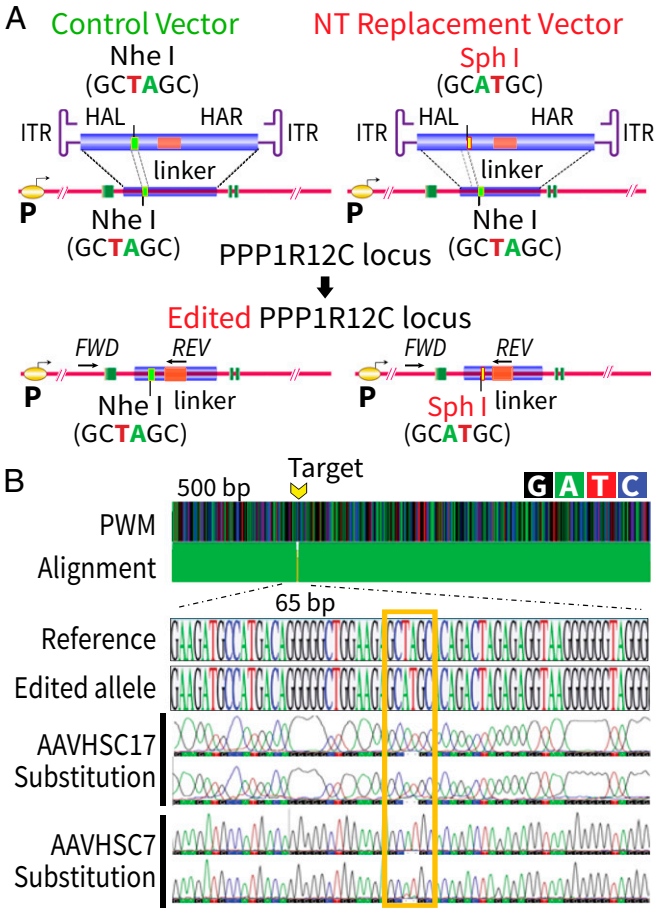

Fig. 3. Replacement of a specific nucleotide to model correction of point mutations. (*A*) Schema of dinucleotide replacement editing assay. (*Right*) In line 1, the nucleotide (NT) replacement vector was designed to replace the wild-type genomic "TA" in intron 1 of PPP1R12C with "AT", replacing the wild-type NheI site with a SphI site. The nucleotide replacement vector was also designed to insert a 12-bp linker 245 bp downstream (depicted in orange). (*Left*) In line 1, control vector was designed to only insert the 12-bp linker, without replacing the TA dinucleotide. HAL, left HA; HAR, right HA. Line 2 shows the wild-type chromosome, highlighting the region to be edited. Line 3 shows the expected configuration of the edited chromosome. Green boxes represent exons, and introns are shown as pink lines. *FWD*, forward; *REV*, reverse. (*B*) Sequence alignment showing dinucleotide editing in intron 1 of PPP1R12C in HepG2 cells. Both 500-bp and 80-bp windows are shown. Alignment across sequenced regions is displayed as a stack plot by position (green). Precise dinucleotide editing is displayed as a cumulative position-weighted matrix (PWM) of each edited sequence aligned to a reference trace.

Table S1). CD34$^+$, K562, and HepG2 cells were treated with either the nucleotide replacement (SphI) or control (NheI) vector packaged in AAVHSC7 and AAVHSC17. A 1,002-bp PCR amplicon was generated using a primer specific for the 12-bp linker and a chromosome-specific primer located upstream of the left HA (Fig. 3*A*). Sanger sequencing of amplicons from transduced HepG2 cells was compared with the reference control genomic sequence (Fig. 3*B*). The dinucleotide sequence change was found to be introduced precisely at the correct location, with no additional sequence alterations observed either locally (500-bp level) or at the nucleotide level (Fig. 3*B*). These data demonstrate that AAVHSCs can precisely replace specified genomic nucleotides within the genome without any other concurrent sequence alterations. These data also demonstrate that AAV editing vector genomes as small as 2011b effectively mediate genome editing. Self-complementary AAV (scAAV) editing vectors were not tested for nucleotide replacement since previous reports showed a complete loss of detectable editing with scAAV (26, 46). We conclude that AAVHSC editing vectors may be used for high-fidelity correction of point mutations or for the precise editing of nucleotide variants without the introduction of inadvertent base changes or genomic scarring.

Genome Editing by AAVHSC Requires BRCA2. To identify the mechanism of AAVHSC editing, we tested the hypothesis that mutations in specific genes known to mediate DNA damage response and maintain genome integrity would result in altered editing outcomes (47–59). Targeted insertion of the promoterless GFP ORF into PPP1R12C was quantitated in each of 12 patient-derived B LCLs that harbor loss-of-function mutations in selected DNA repair genes and compared with that in CD34$^+$ cells (Fig. 4*A* and *SI Appendix*, Table S7). Cumulative averages of editing for each AAV serotype across mutant cell lines revealed a spectrum of impacts (49–59) (Fig. 4*B* and *SI Appendix*, Fig. S9*A*). Cell lines with certain mutations displayed potentiated HR, while others displayed diminished HR (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S9). However, since the LCLs were derived from different patients and therefore not congenic, this precluded precise conclusions on the role of each repair gene. Interestingly, however, the overall pattern of each AAV serotype was similar for each mutation, suggesting that the impact of specific repair genes on genome editing was similar for each AAV serotype (Fig. 4*B*).

AAVHSC vectors successfully mediated editing of every DNA repair mutant line tested with the single exception of the cell line GM13023 (59), a compound heterozygous mutant of BRCA2 (*SI Appendix*, Table S7), an essential mediator of HR (60, 61) (Fig. 4 *A* and *B*), indicating an absolute requirement for functional BRCA2 for AAVHSC editing. Cells that were heterozygous for functional BRCA2 (GM14622; *SI Appendix*, Table S7), however, underwent efficient editing, suggesting that a single functional BRCA2 allele was sufficient (Fig. 4 *A* and *B*). The absolute requirement for BRCA2 was further confirmed using the unrelated BRCA2-null fibroblast line EUFA423 (62) (Fig. 4*C*). BRCA2-null fibroblasts were transduced with either a control promoter-driven AAVHSC17 GFP gene transfer vector or a promoterless AAVHSC17 PPP1R12C-GFP editing vector. Flow analysis revealed the complete lack of GFP expression in BRCA2-null cells treated with the editing vector despite robust GFP expression following transduction with the control gene transfer vector (Fig. 4*C*). The difference in the level of GFP expression in the BRCA2-null versus wild-type fibroblasts following transduction with the gene transfer vector is likely due to differences in background genes, since the two fibroblast cell lines were derived from unrelated individuals. These results clearly demonstrate that AAVHSC17-mediated genome editing did not occur in the absence of BRCA2, despite the ability of the vector to efficiently transduce EUFA423 cells. The same AAVHSC17 PPP1R12C-GFP editing vector readily mediated genome editing in other BRCA2$^{+/+}$ cell lines and primary cells (Figs. 1*C*, 2*A*, and 4*A*). Thus, we conclude that BRCA2 is essential for AAVHSC-mediated genome editing. We additionally confirmed the complete absence of editing at other loci in the human genome, including the FVIII and IL2RG genes in BRCA2$^{-/-}$ cells. These results, together with the complete absence of editing by any AAVHSC vector in BRCA2$^{-/-}$ cell lines (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S9), strongly support the requirement for BRCA2 in the AAVHSC-mediated genome editing pathway and demonstrate that HR is the operative mechanism of genome editing by AAVHSC.

AAVHSC Editing Vectors Mediate in Vivo Editing. To evaluate the potential of AAVHSC for in vivo genome editing, we tested whether i.v. injection of an AAVHSC editing vector could efficiently edit a murine genomic locus. The editing vector, Rosa26-Luc, was designed to insert a promoterless luciferase ORF into intron 1 of the ubiquitously expressed murine Rosa26 locus (63, 64) (Fig. 5*A* and *SI Appendix*, Table S1), leading to expression from the endogenous Rosa26 promoter. The promoterless luciferase ORF was cloned downstream of a SA/T2A sequence and

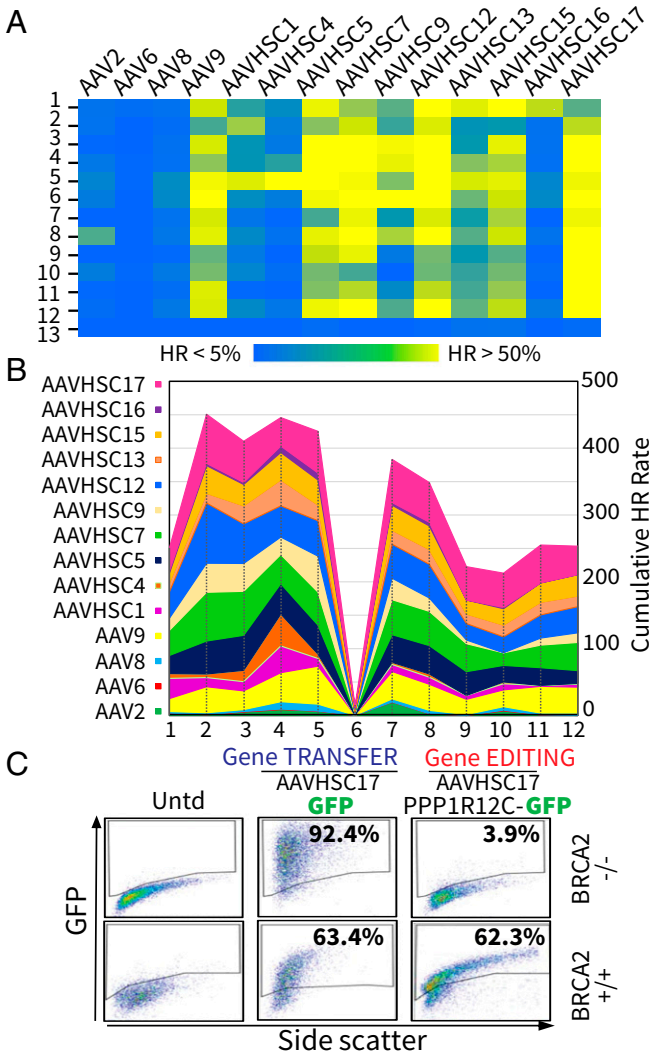

Fig. 4. Role of DNA repair genes in AAVHSC- and AAV9-mediated genome editing. (*A*) Heat map of genome editing efficiencies in LCLs harboring mutations in DNA repair genes. LCLs were treated with AAVHSC PPP1R12C-GFP vectors at a MOI of 1.5E5. Cells were analyzed by flow cytometry after 48 h. Data represent percent GFP expression in live cells following editing of the GFP ORF into intron 1 of PPP1R12C (Fig. 1*A*). Row 1 depicts primary human $CD34^+$ cells for comparison. Subsequent rows depict immortalized cell lines bearing DNA repair mutations as follows: row 2, BLM; row 3, ERCC4/XPF; row 4, NBS1; row 5, RAG1; row 6, ATM; row 7, FANCF; row 8, FANCB; row 9, FANCC; row 10, FANCA; row 11, FANCD2; row 12, $BRCA2^{+/-}$; and row 13, $BRCA2^{-/-}$. Columns depict different AAV serotypes. (*B*) Stack plot of average cumulative editing efficiencies in LCLs harboring loss-of-function mutations in selected DNA damage response genes. Genes in the *x* axis are as follows: 1, GM04408 (BLM); 2, GM08437 (ERCC4); 3, GM15818 (NBS1); 4, ID00078 (RAG1); 5, GM03332 (ATM); 6, GM13023 ($BRCA2^{-/-}$); 7, GM13071 (FANCB); 8, GM14622 ($BRCA2^{+/-}$); 9, GM12794 (FANCC); 10, GM16749 (FANCA); 11, GM16756 (FANCD2); and 12, GM16757 (FANCF). Each AAV serotype is denoted by a specific color. (*C*) AAVHSCs successfully mediate gene transfer into $BRCA2^{-/-}$ fibroblasts but do not edit. Flow cytometric analysis of the EUFA423 $BRCA2^{-/-}$ or $BRCA2^{+/+}$ fibroblast line transduced with AAVHSC17 PPP1R12C-GFP editing vector or a CBA-GFP gene transfer vector is shown, where GFP expression is driven by the CBA promoter. Robust GFP expression is observed after gene transfer, but no GFP is observed following treatment with the editing vector in $BRCA2^{-/-}$ fibroblasts. The absence of editing in $BRCA2^{-/-}$ cells is denoted by lack of GFP expression. Untd, untreated.

was followed by a pA signal (Fig. 5*A*). The entire cassette was flanked by 800-bp HAs (Fig. 5*A* and *SI Appendix*, Table S1). The editing vector, Rosa26-Luc, and a control vector containing the luciferase cassette without HAs (noHA) were packaged in AAVHSC15. Nonobese diabetic/SCID mice were injected i.v. with 5e11 vg with the AAVHSC15 Rosa26-Luc editing vector, the AAVHSC15 noHA control vector, or an AAV8 Rosa26-Luc editing vector. AAVHSC15 and AAV8 were chosen for these experiments since both serotypes have previously been shown to mediate efficient in vivo gene transfer in mice (33, 65, 66) (*SI Appendix*, Fig. S10*A*). Serial whole-body in vivo bioluminescent imaging revealed luciferase expression only in mice injected with the promoterless AAVHSC15 Rosa26-Luc editing vector, but not in mice injected with the control noHA vector lacking HAs, and only nominal expression in mice injected with the AAV8 Rosa26-Luc editing vector (Fig. 5*B* and *SI Appendix*, Figs. S10 and S11). The complete absence of luciferase expression in mice treated with noHA vectors confirmed the requirement for HAs for targeting.

We compared the ability of AAVHSC15 and AAV8 to mediate gene transfer and genome editing in vivo (*SI Appendix*, Fig. S10). Mice were injected with either a gene transfer vector encoding the luciferase gene under control of the chicken beta actin (CBA) promoter or the promoterless Rosa26-Luc editing vector packaged in either AAV8 or AAVHSC15. Bioluminescent imaging and flux measurements 50–60 d postinjection revealed that AAVHSC15 mediated in vivo gene transfer 3.7-fold more efficiently than AAV8 (*SI Appendix*, Fig. S10*A*). However, for mice injected with the promoterless Rosa26-Luc editing vector, the AAVHSC15-treated group showed 43.3-fold greater luciferase expression compared with the AAV8-treated group (*SI Appendix*, Fig. S10*B*). The low level of editing-specific luciferase expression observed in the AAV8 Rosa26-Luc–treated group is comparable to previously described results (27, 67, 68). The markedly increased luciferase expression observed with AAVHSC15 relative to AAV8 was consistent with the enhanced editing capacity of AAVHSCs relative to other AAV vectors in vitro (Figs. 1*C* and 5*B* and *SI Appendix*, Figs. S1 and S2).

Luciferase expression was detectable as early as day 3 after injection with the AAVHSC15 Rosa26-Luc editing vector, and was stable to day 112 postinjection, the last time point assayed. Expression gradually increased after injection and plateaued within ~4–6 wk. Luciferase expression was observed systemically, consistent with the expected ubiquitous expression of Rosa26 (69). In vivo imaging indicated strong widespread systemic luciferase expression (Fig. 5*B*). Organ-specific expression was assessed in isolated organs at the end of the experiment. Quantitation of flux in isolated organs revealed the highest luciferase expression, as measured by flux, in the liver, followed by muscle (*SI Appendix*, Table S8). Luciferase expression was also detected in the heart, lungs, kidney, and brain. Quantitation of vector was performed for isolated organs by ddPCR specific for the luciferase gene relative to a single-copy endogenous gene, apoB. The liver showed the most copies of the luciferase gene at 0.737 copies per cell, followed by muscle (0.398 copies per cell) and heart (0.317 copies per cell) (*SI Appendix*, Table S8). Notably, no toxicity due to AAVHSC editing was noted for up to 6 mo postinjection, the end of the study.

To confirm editing at the molecular level, we employed linear amplification PCR (LAM-PCR) followed by sequence analyses. LAM-PCR was initiated in the chromosomal sequences external to the HAs, spanning both HAs and reading into the luciferase ORF. Sequence analyses confirmed accurate insertion into the intended location, intron 1 of the Rosa26 gene (Fig. 5*C*). No indel mutations or AAV ITRs were detected in the edited genomes. We further confirmed TI of the luciferase cassette into the Rosa26 locus by Southern blot analysis of mouse liver DNA, harvested 70 d after injection of AAVHSC15 Rosa26-Luc. The expected 7-kb band was visualized following hybridization of SpeI-digested DNA with a luciferase-specific probe (Fig. 5*D*). No luciferase-specific bands were detectable in untreated liver DNA from control mice (Fig. 5*D*), showing specificity. Interestingly, no 3,966-bp monomer-length free vector genomes were detected in the Southern blot analysis, possibly suggesting that the editing genome may either not have initiated second-strand synthesis or not formed double-stranded concatamers commonly observed

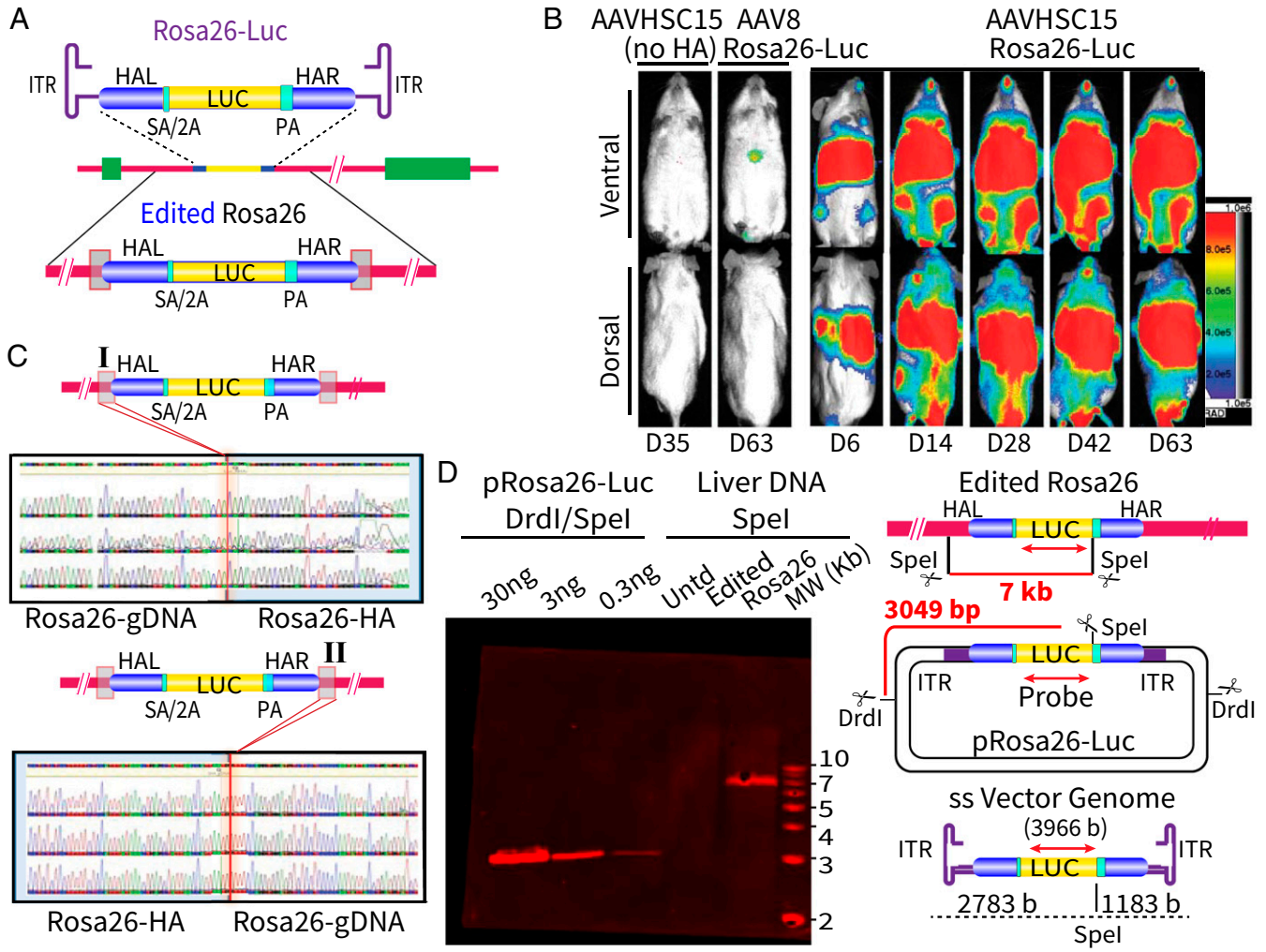

Fig. 5. AAVHSCs mediate efficient in vivo genome editing after i.v. injection. (*A*) Schema showing the vector and genome maps for the Rosa26-Luc editing vector that inserts the promoterless luciferase ORF into intron 1 of the murine Rosa26 locus. The luciferase ORF is preceded by SA/T2A (SA/2A) and followed by pA (PA). The luciferase cassette is flanked by 800-bp HAs targeting intron 1 of Rosa26. HAL, left HA; HAR, right HA. Also depicted is the map of the edited Rosa26 locus. (*B*) Serial in vivo imaging of luciferase expression in mice injected with AAVHSC15 Rosa26-Luc vector (5e11 vg per mouse). Luciferase expression is under the control of the chromosomal Rosa26 promoter. Also shown are AAVHSC15 noHA, a negative control containing the luciferase cassette but noHA, and an AAV8 Rosa26-Luc editing vector. Days postinjection (D) are depicted below each image. The ventral (*Upper*) and dorsal (*Lower*) images of each mouse are shown. Sample sizes for the experimental groups are as follows: AAVHSC15 Rosa26-Luc group, $n = 5$; AAVHSC15 noHA group, $n = 3$; and AAV8 Rosa26-Luc group, $n = 3$. (*C*) Sanger sequence analysis of the junction sequences between the HA and genomic DNA (gDNA) confirmed precise editing of the luciferase ORF into the Rosa26 locus. (*D*) Southern blot analysis of the edited murine Rosa26 locus. Genomic DNA from the liver was digested with SpeI, gel-electrophoresed, transferred to a nylon membrane, and hybridized with a luciferase-specific probe. The expected 7-kb band is observed with the luciferase probe in the edited liver DNA (lane 5), but not from untreated liver DNA (lane 4). Also shown is a titration of the vector plasmid as a positive control for the probe (lanes 1–3). MW, molecular weight marker. Maps show the location of the 7-kb band in the edited Rosa26 locus and the 3,049-bp band in the Rosa26-Luc vector plasmid (pRosa26-Luc) cut with DrdI and SpeI. Potential sizes are also depicted on a map of the vector genome. Red double-headed arrows depict the luciferase-binding site.

following AAV transduction (70). Digestion of double-stranded concatamers with a single cutter enzyme would be expected to yield monomer-length AAV vector genomes, which were not observed, suggesting that such concatamers were absent. If any single-stranded AAV genomes were present, they would likely have formed "snapback" structures due to the complementarity of the ITRs and may have migrated faster in nondenaturing gels. If free vector genomes persisted as double-stranded monomers, SpeI digestion would yield fragments of 2,783 bp and 1,183 bp (Fig. 5*D*), which were not observed. It is also possible that the number of free vector genomes was below the limit of detection for the Southern blot assay (71). However, the presence of a 7-kb band on the Southern blot analysis, together with in vivo expression and sequence analysis, confirmed the accurate targeted insertion of the luciferase ORF into intron 1 of Rosa26 following a single i.v. injection of the AAVHSC15 editing vector. Southern blot analysis also revealed no discernible off-target bands, suggesting the absence of off-target integrations within the limits of detection of the assay.

Discussion

Here, we present evidence for a unique, precise, exclusively HR-based platform for genome editing that operates at unprecedented efficiencies in primary cells in vitro and in vivo, without the need for prior nuclease-mediated DNA breaks. The genomic site to be edited is specified solely by the HAs. Remarkably, this editing platform functions efficiently in postmitotic cells, such as the adult liver in vivo, freshly isolated human liver sections, myocytes, and unstimulated primary $CD34^+$ cells. On-target editing events were highly precise and seamless, with no evidence of indels. This editing pathway functions exclusively through HR, as evidenced by the absolute requirement for BRCA2, an essential mediator of HR, and the complete absence of any evidence of NHEJ-based

repair, including indels and incorporation of AAV ITRs. This genome editing modality may be used not only for efficient editing of therapeutic sequences at specified chromosomal locations but also for accurate correction of disease-causing point mutations. While AAV-mediated genome editing in the absence of nucleases has been previously reported (24–29, 67, 68), efficiencies were too low for most therapeutic applications (29). Here, we present data supporting genome editing efficiencies >50% in primary human cells as measured by expression of a reporter gene following targeted chromosomal insertion. This property of enhanced genome editing efficiency is uniquely shared by members of stem cell-derived AAV clade F, including AAVHSCs and AAV9, and provides a distinctive alternative to current genome editing platforms.

Recently, AAV6 has been used in conjunction with nuclease-based genome editing platforms for the delivery of donor DNA for homology-dependent repair of double-stranded DNA breaks (72–76). While AAV6-mediated donor delivery was efficient and editing was observed following the creation of DNA breaks, no AAV6-mediated editing was observed in the absence of nucleases (72–75), in concordance with our findings.

Our results indicate that vector genomes ranging from ∼2 to 4 kb were found to mediate enhanced genome editing, suggesting that genome size may not be critical. The editing vectors used here were designed with 800b symmetrical HAs, with the exception of the nucleotide substitution vector, which had asymmetrical HAs of 583b and 1045b, indicating that flexibility in HA design is feasible.

This distinctive editing platform is based upon the naturally occurring AAVs present in human HSCs, including AAVHSCs and AAV9 (30). The unique single-stranded AAV genomes bounded by the palindromic G/C-rich ITRs likely activate a cellular HR pathway, which leads to genome editing. This is supported by previous observations that scAAV genomes, which are fully double-stranded, do not mediate recombination (26, 46). Results presented here suggest that in addition to the unique AAV genome structure, the capsid sequences of clade F AAV play an essential role in potentiating HR and genome editing. Identical editing vector genomes packaged in non-clade F capsids, including AAV2, AAV6, or AAV8, mediate editing at significantly lower efficiencies both in vitro and in vivo.

Our results indicate that clade F AAVs were over 27-fold more efficient at nuclear entry than AAVs belonging to clades A, B, and E. Since nuclear entry is essential for both genome editing and transduction, it is possible that more efficient nuclear entry might be partially responsible for the higher level of genome editing observed with clade F viruses, potentially accounting for the observed correlation between transduction and editing efficiencies in vitro. However, in vivo comparisons of gene transfer with genome editing via AAVHSC15 and AAV8 suggested that editing efficiency did not directly correlate with transduction efficiency in vivo. Although both AAV8 and AAVHSC15 transduced mice efficiently, AAVHSC15 was more efficient at editing than AAV8. Whether specific HR pathways are additionally initiated upon receptor engagement by AAVHSC or other cellular mechanisms are activated upon interaction of target cells with the capsids is under investigation.

Upon nuclear entry and uncoating, single-stranded AAV genomes undergo second-strand synthesis and often exist as circular concatamers. Interestingly, under certain conditions, such as in the absence of ATM, AAV genomes were observed to remain in a linear configuration, inaccessible to the NHEJ pathway (70). Notably, no free vector genomes were detected in our Southern blot analysis, consistent with the possibility that circular concatamers were not generated. It is conceivable that a prolonged single-stranded phase of editing vector genomes may additionally lead to enhanced editing. While single-stranded vector genomes could potentially be unstable in the nuclear environment, association with a cascade of repair proteins may serve to protect them from degradation and promote HR.

Importantly, the ability of AAVHSC and AAV9 to mediate HR efficiently in resting and postmitotic cells was evidenced by successful editing of freshly isolated liver sections, myocytes, and unstimulated $CD34^+$ cells, as well as in vivo, including the liver and muscle in adult mice. The ability of AAVHSC and AAV9 to support genome editing in postmitotic adult tissue is unique (17) and has both mechanistic as well as therapeutic implications.

Thus, AAV clade F, including AAVHSC and AAV9, represents a uniquely HR-based, highly precise, and efficient seamless genome editing modality in a single-component, nuclease-free platform, with a built-in delivery component that overcomes many of the challenges currently facing therapeutic translation of genome editing. Thus, together with its potent in vivo editing ability, clade F AAVs represent a distinct platform for the development of therapeutic in vivo genome editing for the treatment and potential cure of human diseases.

Materials and Methods

All animal care and experiments were performed under protocols approved by the City of Hope Institutional Animal Care and Use Committee. Deidentified, cytokine-primed, peripheral blood $CD34^+$ cells were obtained with informed consent from healthy donors under a City of Hope Institutional Review Board (IRB)-approved protocol as previously described (29). Deidentified liver samples were obtained from the City of Hope operating room under an IRB-approved protocol for the use of discard material. Recombinant DNA work was performed according to NIH guidelines. Tissue culture, transductions, flow cytometry, Southern blot analysis, and PCR reactions were performed using standard procedures unless otherwise specified. A more detailed description of treatments and analyses is provided in *SI Appendix, SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Dr. Indra M. Newman for valuable advice and assistance with the manuscript; Drs. Stephen Elledge, Kush Parmar, Albert Seymour, Arthur Tzianabos, and Sam Rasty for critical scientific feedback; and Drs. James McSwiggen and Mike O'Callaghan for stimulating discussions, scientific feedback, and helpful interactions. City of Hope core facilities utilized for this study include the Analytical Cytometry, DNA Sequencing, and Small Animal Imaging Cores. This research was supported by the National Heart, Lung, and Blood Institute and National Cancer Institute of the NIH through Grant HL087285 (to S.C.) and by Cancer Center Support Grant P30CA033572.

1. Jasin M (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* 12:224–228.
2. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* 33:25–35.
3. Pâques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* 63:349–404.
4. Bibikova M, Golic M, Golic KG, Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* 161:1169–1175.
5. Zhang F, et al. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 29:149–153.
6. Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29:143–148.
7. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339: 823–826.
8. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823.
9. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
10. Zetsche B, et al. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:759–771.
11. Lieber MR (2010) The mechanism of double-strand DNA break repair by the non-homologous DNA end-joining pathway. *Annu Rev Biochem* 79:181–211.
12. Jeggo PA (1998) DNA breakage and repair. *Adv Genet* 38:185–218.
13. Daya S, Cortez N, Berns KI (2009) Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway. *J Virol* 83:11655–11664.
14. Rudin N, Haber JE (1988) Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. *Mol Cell Biol* 8:3918–3928.
15. Capecchi MR (1989) Altering the genome by homologous recombination. *Science* 244: 1288–1292.
16. Rouet P, Smih F, Jasin M (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci USA* 91:6064–6068.
17. Hustedt N, Durocher D (2016) The control of DNA repair by the cell cycle. *Nat Cell Biol* 19:1–9.
18. Cox DB, Platt RJ, Zhang F (2015) Therapeutic genome editing: Prospects and challenges. *Nat Med* 21:121–131.

19. Fu Y, et al. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31:822–826.
20. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* 32: 677–683.
21. Tsai SQ, et al. (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* 33:187–197.
22. Ran FA, et al. (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520:186–191.
23. Komor AC, Badran AH, Liu DR (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* 169:559, erratum (2017) 168:20–36.
24. Russell DW, Hirata RK (1998) Human gene targeting by viral vectors. *Nat Genet* 18: 325–330.
25. Porteus MH, Cathomen T, Weitzman MD, Baltimore D (2003) Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol* 23: 3558–3565.
26. Vasileva A, Linden RM, Jessberger R (2006) Homologous recombination is required for AAV-mediated gene targeting. *Nucleic Acids Res* 34:3345–3360.
27. Barzel A, et al. (2015) Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* 517:360–364.
28. Miller DG, et al. (2006) Gene targeting in vivo by adeno-associated virus vectors. *Nat Biotechnol* 24:1022–1026.
29. Khan IF, Hirata RK, Russell DW (2011) AAV-mediated gene targeting methods for human cells. *Nat Protoc* 6:482–501.
30. Smith LJ, et al. (2014) Gene transfer properties and structural modeling of human stem cell-derived AAV. *Mol Ther* 22:1625–1634.
31. Podsakoff G, Wong KK, Jr, Chatterjee S (1994) Efficient gene transfer into non-dividing cells by adeno-associated virus-based vectors. *J Virol* 68:5656–5666.
32. Santat L, et al. (2005) Recombinant AAV2 transduction of primitive human hematopoietic stem cells capable of serial engraftment in immune-deficient mice. *Proc Natl Acad Sci USA* 102:11053–11058.
33. Zincarelli C, Soltys S, Rengo G, Rabinowitz JE (2008) Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther* 16: 1073–1080.
34. Chatterjee S, Johnson PR, Wong KK, Jr (1992) Dual-target inhibition of HIV-1 in vitro by means of an adeno-associated virus antisense vector. *Science* 258:1485–1488.
35. Fisher-Adams G, Wong KK, Jr, Podsakoff G, Forman SJ, Chatterjee S (1996) Integration of adeno-associated virus vectors in CD34+ human hematopoietic progenitor cells after transduction. *Blood* 88:492–504.
36. Paz H, et al. (2007) Quiescent subpopulations of human CD34-positive hematopoietic stem cells are preferred targets for stable recombinant adeno-associated virus type 2 transduction. *Hum Gene Ther* 18:614–626.
37. Kotin RM, et al. (1990) Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci USA* 87:2211–2215.
38. Zhong L, et al. (2004) Impaired nuclear transport and uncoating limit recombinant adeno-associated virus 2 vector-mediated transduction of primary murine hematopoietic cells. *Hum Gene Ther* 15:1207–1218.
39. Ponnazhagan S, et al. (1997) Adeno-associated virus type 2-mediated transduction in primary human bone marrow-derived CD34+ hematopoietic progenitor cells: Donor variation and correlation of transgene expression with cellular differentiation. *J Virol* 71:8262–8267.
40. Mroske C, Rivera H, Ul-Hasan T, Chatterjee S, Wong KK (2012) A capillary electrophoresis sequencing method for the identification of mutations in the inverted terminal repeats of adeno-associated virus. *Hum Gene Ther Methods* 23:128–136.
41. Sargent RG, Brenneman MA, Wilson JH (1997) Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol* 17:267–277.
42. Pipiras E, Coquelle A, Bieth A, Debatisse M (1998) Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. *EMBO J* 17:325–333.
43. Sherry ST, et al. (2001) dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res* 29:308–311.
44. Genovese P, et al. (2014) Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* 510:235–240.
45. Miyaoka Y, et al. (2016) Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* 6:23549.
46. Hirata RK, Russell DW (2000) Design and packaging of adeno-associated virus gene targeting vectors. *J Virol* 74:4612–4620.
47. Powell SN, Kachnic LA (2003) Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* 22:5784–5791.
48. Oettinger MA, Schatz DG, Gorka C, Baltimore D (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248:1517–1523.
49. Liu P, et al. (1993) Regional mapping of human DNA excision repair gene ERCC4 to chromosome 16p13.13-p13.2. *Mutagenesis* 8:199–205.
50. Meyn MS (1995) Ataxia-telangiectasia and cellular responses to DNA damage. *Cancer Res* 55:5991–6001.
51. Tauchi H, et al. (2002) Nbs1 is essential for DNA repair by homologous recombination in higher vertebrate cells. *Nature* 420:93–98.
52. Ellis NA, et al. (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 83:655–666.
53. Joenje H, et al. (1997) Evidence for at least eight Fanconi anemia genes. *Am J Hum Genet* 61:940–944.
54. Joenje H, et al. (2000) Complementation analysis in Fanconi anemia: Assignment of the reference FA-H patient to group A. *Am J Hum Genet* 67:759–762.
55. Meetei AR, et al. (2004) X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* 36:1219–1224.
56. Ikeda H, et al. (2003) Genetic reversion in an acute myelogenous leukemia cell line from a Fanconi anemia patient with biallelic mutations in BRCA2. *Cancer Res* 63: 2688–2694.
57. Takata M, et al. (2006) The Fanconi anemia pathway promotes homologous recombination repair in DT40 cell line. *Subcell Biochem* 40:295–311.
58. Michl J, Zimmer J, Tarsounas M (2016) Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *EMBO J* 35:909–923.
59. Strathdee CA, Gavish H, Shannon WR, Buchwald M (1992) Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* 356:763–767.
60. Moynahan ME, Pierce AJ, Jasin M (2001) BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol Cell* 7:263–272.
61. Davies AA, et al. (2001) Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol Cell* 7:273–282.
62. Feng Z, et al. (2011) Rad52 inactivation is synthetically lethal with BRCA2 deficiency. *Proc Natl Acad Sci USA* 108:686–691.
63. Friedrich G, Soriano P (1991) Promoter traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. *Genes Dev* 5:1513–1523.
64. Kisseberth WC, Brettingen NT, Lohse JK, Sandgren EP (1999) Ubiquitous expression of marker transgenes in mice and rats. *Dev Biol* 214:128–138.
65. Wang L, et al. (2015) Comparative study of liver gene transfer with AAV vectors based on natural and engineered AAV capsids. *Mol Ther* 23:1877–1887.
66. Smith L, et al. (2011) Functional mapping of tissue tropism of naturally occurring adeno-associated virus isolates from human hematopoietic stem cells. *Mol Ther* 19(Suppl 1):S127–S128.
67. Paulk NK, et al. (2010) Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. *Hepatology* 51:1200–1208.
68. Borel F, et al. (2017) Survival advantage of both human hepatocyte xenografts and genome-edited hepatocytes for treatment of alpha-1 antitrypsin deficiency. *Mol Ther* 25:2477–2489.
69. Giel-Moloney M, Krause DS, Chen G, Van Etten RA, Leiter AB (2007) Ubiquitous and uniform in vivo fluorescence in ROSA26-EGFP BAC transgenic mice. *Genesis* 45:83–89.
70. Cataldi MP, McCarty DM (2010) Differential effects of DNA double-strand break repair pathways on single-strand and self-complementary adeno-associated virus vector genomes. *J Virol* 84:8673–8682.
71. Brown T (1993) Southern blotting. *Curr Protoc Mol Biol* 21:2.9.1–2.9.20.
72. Sather BD, et al. (2015) Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci Transl Med* 7: 307ra156.
73. Wang J, et al. (2015) Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat Biotechnol* 33:1256–1263.
74. Wang J, et al. (2016) Highly efficient homology-driven genome editing in human T cells by combining zinc-finger nuclease mRNA and AAV6 donor delivery. *Nucleic Acids Res* 44:e30.
75. De Ravin SS, et al. (2016) Targeted gene addition in human CD34(+) hematopoietic cells for correction of X-linked chronic granulomatous disease. *Nat Biotechnol* 34: 424–429.
76. Ling C, et al. (2016) High-efficiency transduction of primary human hematopoietic stem/progenitor cells by AAV6 vectors: Strategies for overcoming donor-variation and implications in genome editing. *Sci Rep* 6:35495.