Molecular Therapy
Nucleic Acids
Original Article

Gene Editing with Helper-Dependent Adenovirus Can Efficiently Introduce Multiple Changes Simultaneously over a Large Genomic Region

Donna J. Palmer,[1] Nathan C. Grove,[1] Dustin L. Turner,[1] and Philip Ng[1]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Helper-dependent adenoviral vectors (HDAds) possess long homology arms that mediate high-efficiency gene editing. These long homology arms may permit simultaneous introduction of multiple modifications into a large genomic region or may permit a single HDAd to correct many different individual mutations spread widely across a gene. We investigated this important potential using an HDAd bearing 13 genetic markers in the region of homology to the target *CFTR* locus in human iPSCs and found that all markers can be simultaneously introduced into the target locus, with the two farthest markers being 22.2 kb apart. We found that genetic markers closer to the HDAd's selectable marker are more efficiency introduced into the target locus; a marker located 208 bp from the selectable marker was introduced with 100% efficiency. However, even markers 11 kb from the selectable marker were introduced at a relatively high frequency of 21.7%. Our study also revealed extensive heteroduplex DNA formation of up to 10 kb with no bias toward vector or chromosomal repair. However, mismatches escape repair at a frequency of up to 15%, leading to a genetically mixed colony and emphasizing the need for caution, especially if the donor and target sequences are not 100% homologous.

INTRODUCTION

We and others have shown that helper-dependent adenoviral vectors (HDAds) can efficiently deliver donor DNA into human embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs), and adult stem cells to achieve high-efficiency gene editing by spontaneous homologous recombination (HR) to a variety of transcriptionally active and inactive loci and thus to achieve knockins, knockouts, and corrections.[1–13] Altogether, these studies have consistently demonstrated that HDAd-mediated gene editing of iPSCs and ESCs is not associated with ectopic random HDAd integrations, does not affect the undifferentiated state and pluripotency, and maintains genetic and epigenetic integrity. One study found that targeted gene correction in iPSCs by HDAd minimally affects whole-genome mutational load, as determined by whole genome sequencing,[9] and we have demonstrated that footprintless gene editing can be efficiently achieved in human iPSCs with HDAd.[13]

HDAds are appealing as a gene editing vector for many reasons. A major appeal is that induction of an artificial double-stranded break (DSB) at the chromosomal target locus by a designer endonuclease is not required to achieve high targeting efficiency, thereby eliminating the potential for off-target cleavage. In addition, HDAds can efficiently deliver foreign DNA into the nucleus of target cells. Furthermore, they are deleted of all viral-coding sequences, thereby reducing their toxicity and increasing their cloning capacity to 37 kb. This tremendous cloning capacity permits inclusion of long homology arms, multiple selectable markers, promoters, enhancers, *cis*-acting elements, and other transgenes to enhance their gene editing efficiency. In addition, the adenoviral terminal protein is covalently attached to the ends of the linear double-stranded DNA genome of the HDAd, and this minimizes promiscuous interactions between the HDAd genome and the host chromosomes, greatly reducing random vector integration.[14] Finally, HDAds offer the potential for in vivo gene editing.

As mentioned, HDAds can accommodate long homology arms, which we have shown to enhance the efficiency of gene editing.[13] In addition, Liu et al.[4] showed that a single HDAd could correct two different, albeit closely spaced, mutations at the lamin A (LMNA) locus in human iPSCs. Subsequent analysis revealed that a single nucleotide polymorphism (SNP) in the HDAd homology arm located 4.4 kb from the positive selectable marker was also introduced into the target locus. This led Liu et al.[4] to suggest that a single HDAd could be used to replace a 4.4 kb region bearing exons 3 to 12 and thus a single HDAd could be used to correct the more than 200 laminopathy-associated LMNA mutations that have been described in these exons.[4] A corollary put forth by Li et al.[15] is that long homology arms may offer the potential to simultaneously introduce multiple targeted modifications into a large region of chromosomal DNA. In this study, we investigate this potentially powerful gene editing property of HDAds to determine its capabilities and limits.

RESULTS

Experimental System

The objective of this study was to determine the potential of the long homology arms in HDAds for introducing modifications along their

Received 10 May 2017; accepted 1 June 2017;
http://dx.doi.org/10.1016/j.omtn.2017.06.001.

Correspondence: Philip Ng, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.
E-mail: png@bcm.edu

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

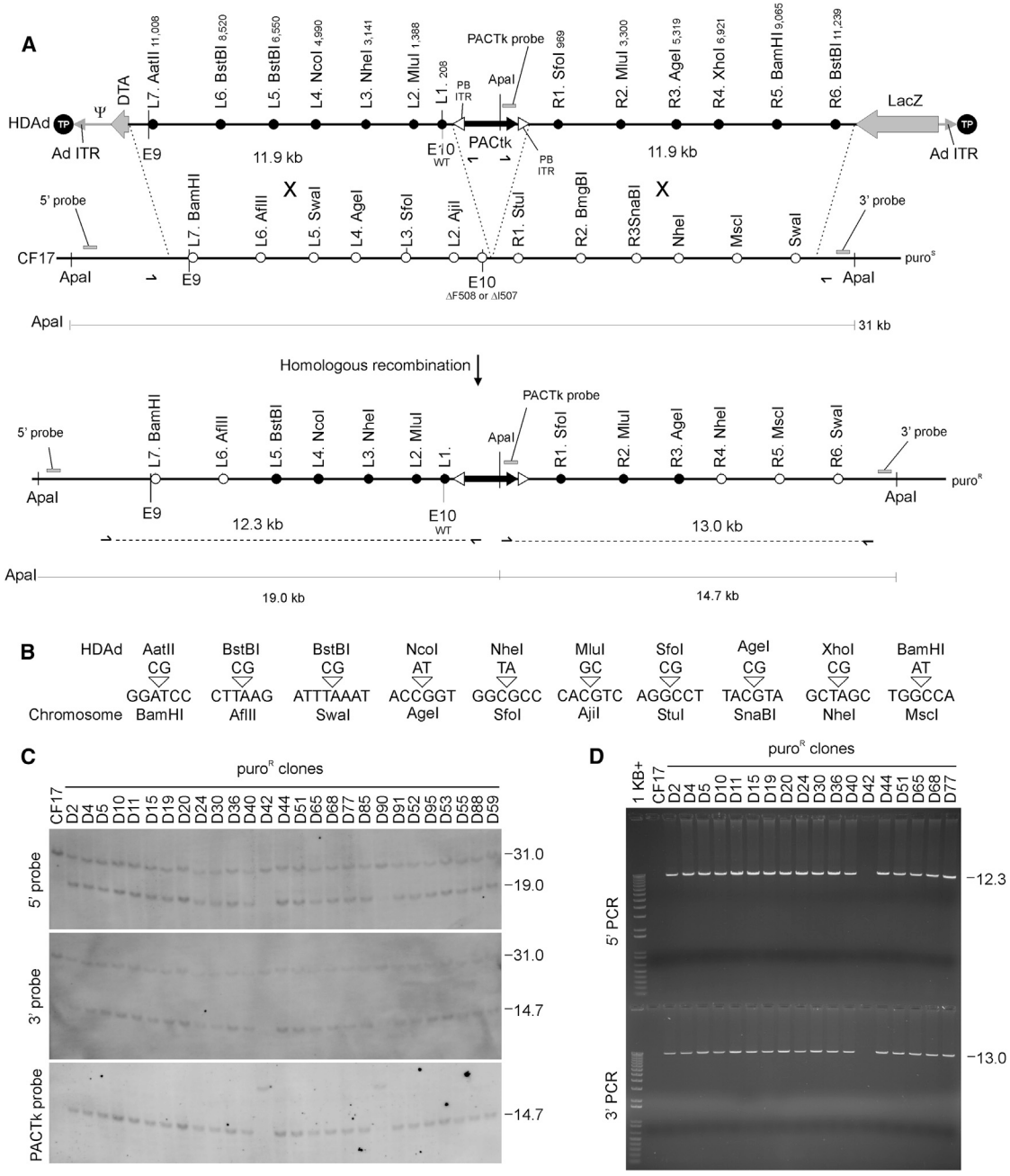

Figure 1. Gene Targeting at the *CFTR* Locus by HDAd

(A) A single reciprocal crossover in the right and left homology arms of the HDAd (called HD-23.8-CFTRm-PACTk-DTA) results in integration of PACTk into the *CFTR* gene rendering recombinants puro^R. The poly(A)-less diphtheria toxin A (DTA) fragment gene in the HDAd permits negative selection against clones bearing random HDAd integrations. PiggyBac inverted terminal repeats (PB ITRs) flank PACTk to permit its footprintless excision in the presence of PB transposase. The sizes of the diagnostic ApaI fragments and the locations of the 5′ external probe, 3′ external probe, and internal PACTk probe used for Southern analyses are shown. The positions of PCR primers used to amplify the 5′ and 3′ regions of homology in the targeted recombinants are shown. The positions of the adenoviral packaging signal (ψ) and adenoviral inverted terminal repeats (Ad ITRs) are shown. Genetic markers that distinguish the vector from the chromosomal sequence in the left homology arm are designated L1 through L7. Genetic markers that distinguish the vector from the chromosomal sequence in the right homology arm are designated R1 through R6. The identities of all genetic markers are shown, as well as their distance in base pairs (bps) from the closest PB ITR. Also shown are the adenoviral terminal proteins (TPs) that are covalently attached to the ends of the linear double-stranded DNA genome of the HDAd. The location of *CFTR* exon 9 (E9) and exon 10 (E10) in the vector and chromosome are also shown. The vector bears the wild-type *CFTR* sequence in E10, whereas the chromosomal E10 contains the ΔF508 mutation in one allele and the ΔI507 mutation in the other allele. The distance between L7

(legend continued on next page)

Table 1. Gene Targeting Frequencies at the CFTR Locus with HDAd

Total Puro^R Colonies[a]	Puro^R Colonies Analyzed by Southern Blot[b]	Correctly Targeted[c]	Aberrantly Targeted[d]	Random Vector Integration[e]
177	95	85 (89.5%)	2 (2.1%)	8 (8.4%)

[a]Total number of puro^R colonies obtained after transduction of 2×10^6 CF17 iPSCs with HD-23.8-CFTRm-PACTk-DTA at an MOI of 350 vp/cell.
[b]Number of puroR colonies analyzed by Southern blot.
[c]Number of correctly targeted clones as determined by Southern blot and confirmed by PCR.
[d]Number of aberrantly targeted clones as determined by Southern blot and confirmed by PCR.
[e]Number of clones with random HDAd integration as determined by Southern blot and confirmed by PCR.

entire length into the chromosomal target locus. As a model system, we targeted the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene in a human iPSC line, called CF17, which is compound heterozygous at the *CFTR* locus, with one allele ΔF508 and the other allele ΔI507 in exon 10.[16] To accomplish this, we started with HD-23.8-CFTR-PACTk, an HDAd that we previously showed to achieve high-efficiency footprintless gene editing at the *CFTR* locus of CF17 iPSCs,[13] and modified it to create a new HDAd called HD-23.8-CFTRm-PACTk-DTA (Figure 1A). The modifications composed of twelve 2 bp insertions, six (markers L2 to L7) in the 11.9 kb left homology arm and six (markers R1 to R6) in the 11.9 kb right homology arm. Each of these twelve 2 bp insertions converts an endogenous restriction enzyme site into a novel one (Figure 1B); thus, the contribution of vector sequences at the target locus following gene targeting can be easily determined as described later. In addition to these twelve markers, HD-23.8-CFTRm-PACTk-DTA contains the wild-type *CFTR* exon 10 in the left homology arm (marker L1), whereas the exon 10 at the chromosomal *CFTR* target locus contains the ΔF508 mutation in one allele and ΔI507 in the other. Finally, HD-23.8-CFTRm-PACTk-DTA, unlike its predecessor, contains the poly(A)-less diphtheria toxin A fragment gene to permit passive negative selection against random vector integrations.[17] Like its predecessor, the homology arms of HD-23.8-CFTRm-PACTk-DTA flank a puromycin *N*-acetyltransferase-HSV-TK fusion gene (PACTk), which permits positive selection for HDAd integration by conferring puromycin resistance (puro^R). The PACTk cassette is flanked by piggyBac inverted terminal repeats (PB ITRs), which we have previously shown to permit footprintless excision of PACTk in the presence of PB transposase, thus conferring resistance to ganciclovir.[13]

HDAd-Mediated Gene Targeting Frequency

To determine the contribution of vector sequence at the targeted locus, 2×10^6 CF17 cells were transduced with HD-23.8-CFTRm-PACTk-DTA at an MOI of 350 viral particles per cell (vp/cell), which resulted in a total of 177 puro^R colonies (Table 1). To determine the frequency of correct gene targeting, DNA was extracted from 95 (of the 177) well-isolated puro^R colonies, digested with ApaI, and subjected to Southern blot analyses. As shown in Figure 1A, targeted vector integration into the *CFTR* locus by two reciprocal crossovers, one in each homology arm, converts the endogenous 31 kb ApaI fragment (revealed by the external 5′ and 3′ probes) into a 19 kb ApaI fragment (revealed by the external 5′ probe) and a 14.7 kb ApaI fragment (revealed by the external 3′ probe and the internal PACTk probe). The results of the Southern analyses are summarized in Table 1, and representative Southern blots are shown in Figure 1C. These analyses revealed that of the 95 clones analyzed, 85 (89.5%) were correctly targeted, 2 (2.1%) were aberrantly targeted (1 with correct targeting only in the 5′ homology arm and the other with correct targeting only in the 3′ homology arm), and 8 (8.9%) had random vector integration. The Southern blots also revealed the continued presence of the 31 kb band when analyzed with the external 5′ and 3′ probes (Figure 1C), indicating that targeted vector integration had occurred in only one of the two *CFTR* alleles in all 85 targeted recombinants, a result consistent with our previous study.[13] The frequency of targeting with HD-23.8-CFTRm-PACTk-DTA was not reduced (and perhaps was even a little higher) compared to HD-23.8-CFTR-PACTk,[13] indicating that the twelve 2 bp insertions in the region of homology did not negatively affect targeting efficiency.

As a first step in determining the identity of the restriction enzyme markers at the targeted locus, PCRs were developed to specifically amplify a unique 12.3 kb fragment encompassing the 5′ region of homology and a unique 13.0 kb fragment encompassing the 3′ region of homology only from the correctly targeted recombinant locus (Figure 1A). As expected, there was complete agreement between the results of the Southern blots and the PCRs: all 85 correctly targeted recombinants identified by Southern analyses yielded the expected 12.3 and 13.0 kb PCR products, the 2 aberrantly targeted clones yielded only the expected 12.3 or 13.0 kb PCR product, and all 8 clones with random vector integration yielded no PCR products. The results of a representative gel showing the 5′ and 3′ PCRs are given in Figure 1D. Thus, these results validate the PCR assays for identification of targeted recombinants at the *CFTR* locus in future studies.

Genetic Marker Analyses in Targeted Recombinants

To determine the identity of the restriction enzyme markers, the 5′ and 3′ PCR products from 83 representative targeted recombinants (out of the 85 total) were digested with the diagnostic restriction enzymes and analyzed by agarose gel electrophoresis. As an example, the left gel in Figure 2 presents the results of this analysis for marker L4 in the 5′ PCR product from 19 targeted recombinants. As shown, the 5′ PCR products amplified from recombinants D2, D5, D10, D11,

and R6 is 25.2 kb in the HDAd or 22.2 kb in the chromosome. (B) Sequences of the genetic markers showing the 2 bp insertions that convert a chromosomal restriction enzyme site into a different restriction enzyme site at the homologous position in the HDAd. (C) Representative Southern blots of genomic DNA extracted from puro^R clones analyzed with the 5′ external probe, the 3′ external probe, and the internal PACTk probe showing targeted vector integration for all clones except D42 and D90, which have randomly integrated into the HDAd genome. (D) Representative agarose gel showing PCR amplifications of the 5′ and 3′ regions of homology from genomic DNA extracted from puro^R clones, confirming targeted vector integration for all clones except D42.

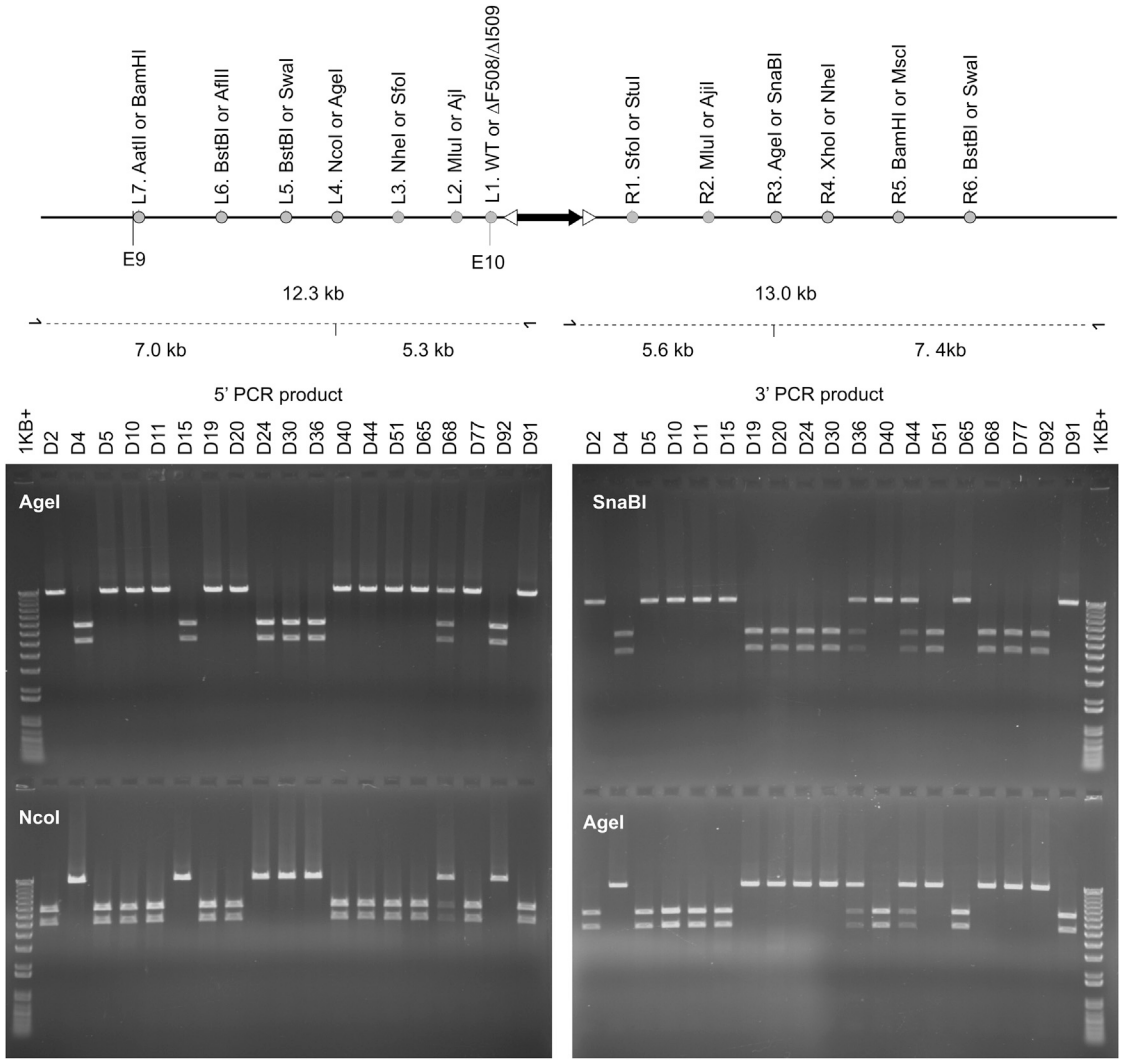

Figure 2. Analyses of Restriction Enzyme Markers

Representative agarose gels of restriction enzyme analyses to determine the identity of the genetic markers in the 5′ and 3′ PCR products. Left gel shows the AgeI and NcoI analyses for the 5′ PCR product to determine the identity of the restriction enzyme site at position L4. Right gel shows the SnaBI and AgeI analyses for the 3′ PCR product to determine the identity of the restriction enzyme site at position R3.

D19, D20, D40, D44, D51, D65, D77, and D91 were resistant to AgeI digestion (as evident by the continued presence of the 12.3 kb fragment) and sensitive to NcoI digestion (as evident by the presence of the 7.0 and 5.3 kb fragments), indicating the presence of the vector-borne NcoI marker in these recombinants. Conversely, the 5′ PCR products amplified from recombinants D4, D15, D24, D30, D36, and D92 were sensitive to AgeI digestion (as evident by the presence of the 7.0 and 5.3 kb fragments) and resistant to NcoI digestion (as evident by the presence of the 12.3 kb fragment), indicating the presence of the chromosomal AgeI marker in these recombinants. The 5′ PCR product from D68 was partially sensitive to both AgeI and NcoI digestion (as evident by the presence of the 12.3, 7.0, and 5.3 kb fragments for both digestions), indicating that recombinant D68 was composed of two subpopulations, one with the vector-borne AgeI marker and the other with the chromosomal NcoI marker. This sectored marker is classic evidence for the formation of heteroduplex DNA (hDNA) encompassing the L4 marker site during HR between the vector and the chromosomal DNAs and that this mismatch escaped mismatch repair (MMR) before DNA replication and cell division, resulting in a mixed (sectored) colony. The gel on the right side of Figure 2 presents an example of the analysis for the marker R3 in the 3′ PCR product from the same 19 targeted recombinants. As shown, the 3′ PCR products amplified from recombinants D2, D5, D10, D11, D15, D40, D65, and D91 were resistant to SnaBI digestion (as evident by the continued presence of the 13.0 kb fragment) and sensitive to AgeI digestion (as evident by the presence of 5.6 and 7.4 kb fragments), indicating the presence of the vector-borne AgeI marker in these recombinants. Conversely, the 3′ PCR products

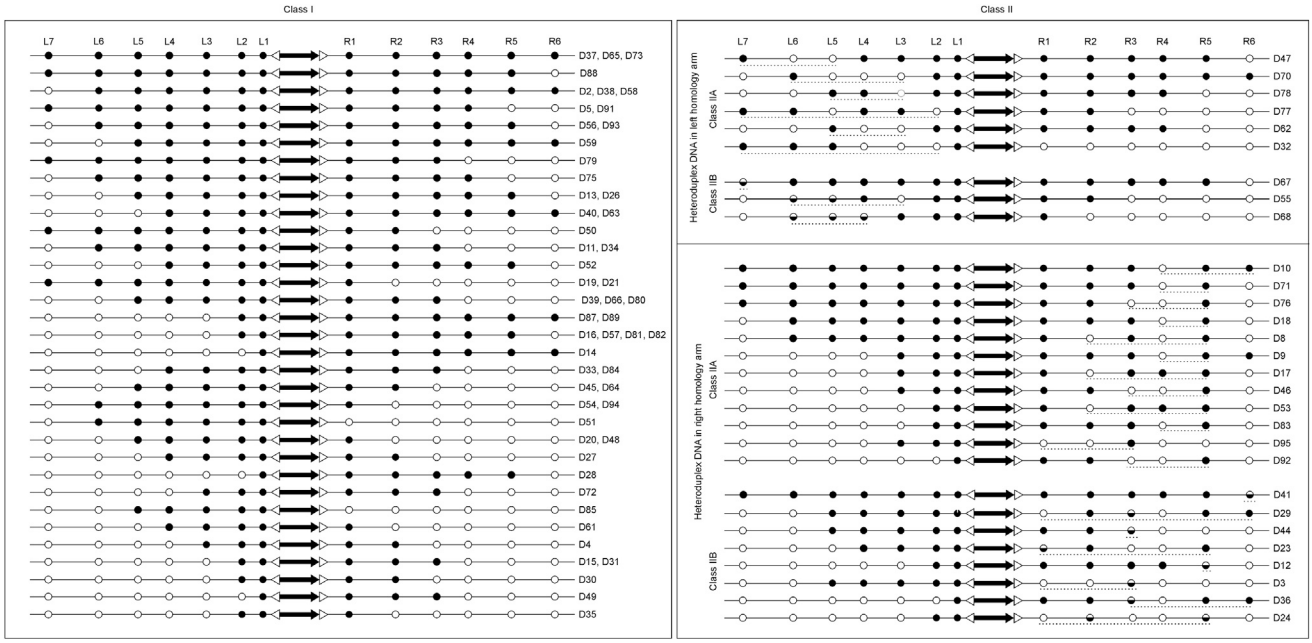

Figure 3. Marker Patterns in Targeted Recombinants

Filled circles represent vector-borne marker, open circles represent chromosomal markers, and half-filled circles represent sectored markers. The linkage relationships between sector markers are as shown. Class I recombinants do not show evidence of hDNA. Class II recombinants show evidence of hDNA, and hDNA tracts are indicated by the broken horizontal lines. Class II recombinants are first separated into those with hDNA in the 5′ or 3′ region of homology. Class II recombinants are further subdivided into subclass A and B; class IIA recombinants have no sectored markers, while class IIB recombinants have sectored markers. Within each class or subclass, the recombinants are arranged in descending order with respect to the number of vector markers present. This organization is arbitrary, because MMR may underestimate the existence, location, and length of hDNA, as explained in the text.

amplified from recombinants D4, D19, D20, D24, D51, D68, D77, and D92 were sensitive to SnaBI digestion (as evident by the presence of the 5.6 and 7.4 kb fragments) and resistant to AgeI digestion (as evident by the presence of the 13.0 kb fragment), indicating the presence of the chromosomal AgeI marker in these recombinants. The 3′ PCR product from recombinants D36 and D44 was partially sensitive to both SnaBI and AgeI digestion (as evident by the presence of the 12.3, 7.0, and 5.3 kb fragments for both digestions), indicating that these recombinants were composed of two subpopulations, one with the vector-borne AgeI marker and the other with the chromosomal SnaBI marker. As explained earlier, this sectored marker indicates that the mismatch formed by hDNA at position R3 escaped MMR. For recombinants whose PCR products possessed more than one sectored marker (D55, D68, and D24), the linkage relationship between the sector markers was determined by double digestions of the PCR product with the involved restriction enzymes. The results revealed that all vector-borne markers were present in one PCR product and, as expected, all chromosomal markers were present on the other PCR product (data not shown). These results are consistent with the sector markers being encompassed within a continuous tract of hDNA. In addition to the 12 restriction enzyme polymorphisms, the vector and chromosome differ in exon 10; whereas the vector contained the wild-type exon 10, the chromosome bears the ΔF508 mutation in one allele and the ΔI507 mutation in the other allele at exon 10. To determine whether the exon 10 at the targeted locus bore the wild-type or mutant (ΔF508 or ΔI507) sequence, the 5′ PCR product was sequenced. The results revealed that all recombinants possessed the wild-type exon 10 (data not shown).

The results of the analyses for all restriction enzyme sites and exon 10 for the 83 recombinants analyzed are shown in Figure 3, with filled circles representing vector-borne markers, open circles representing chromosomal markers, and half-filled circles representing sectored markers. The 83 recombinants were divided into two classes based on their marker patterns. Of the 83 recombinants, 54 were class I, possessing a marker pattern that could have arisen without formation of hDNA during HR. The remaining 29 of the 83 targeted recombinants were class II, possessing a marker pattern consistent with the formation of hDNA during HR. The class II recombinants were divided into two groups; 9 of the 29 class II recombinants had evidence of hDNA in the left homology arm, and 20 of the 29 class II recombinants had evidence of hDNA in the right homology arm. The class II recombinants were further divided into two subclasses, A and B. Class IIA recombinants had mismatches within hDNA repaired before DNA replication, and cell division resulted in pure populations. Class IIB recombinants had mismatches that escaped MMR before DNA replication, and cell division resulted in sectored colonies. For the 9 recombinants with evidence of hDNA in the left homology arm, 6 were class IIA and the remaining 3 were class IIB. For the 20 recombinants with hDNA in the right homology arm, 12 were class IIA and

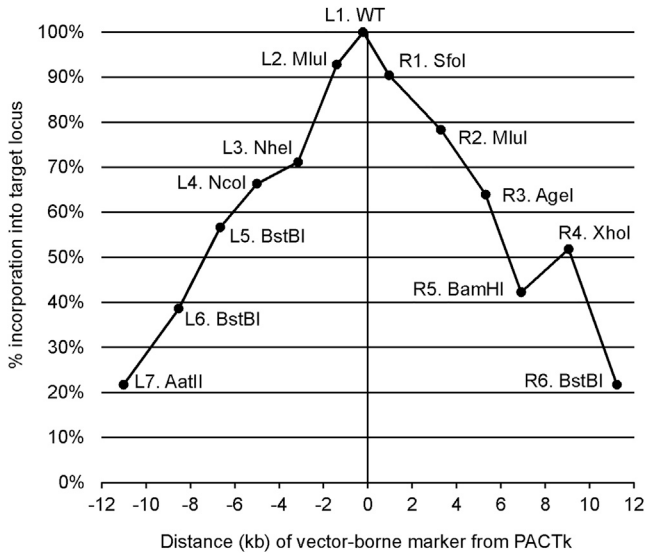

Figure 4. Incorporation Frequencies of Restriction Enzyme Markers

Frequency of incorporation of a vector-borne marker into the recombinant locus as a function of its distance from the selectable PACTk expression cassette following HDAd-mediated gene targeting.

the remaining 8 were class IIB. Finally, the recombinants within each class or subclass are presented in descending order of the number of vector-borne markers at the recombinant locus.

Frequency of Vector-Borne Marker Incorporation into the Target Locus

The marker patterns of the 83 targeted recombinants analyzed revealed a direct correlation between the frequency of incorporation of a vector-borne marker into the recombinant locus and its distance from the PACTk selectable marker (Figure 4). Specifically, the vector-borne wild-type *CFTR* sequence in exon 10, only 208 bp upstream of PACTk, is incorporated 100% (83/83) into the targeted recombinants. The vector-borne MluI marker 1.4 kb upstream of PACTk and the vector-borne SfoI marker 969 bp downstream of PACTk are incorporated into the target locus at frequencies of 92.8% (77/83) and 90.4% (75/83), respectively. The vector-borne NheI marker 5 kb upstream of PACTk and the vector-borne MluI marker 3.3 kb downstream of PACTk are incorporated into the target locus at frequencies of 71.1% (59/83) and 78.3% (65/83), respectively. The vector-borne NcoI marker 5.0 kb upstream of PACTk and the vector-borne AgeI marker 5.3 kb downstream of PACTk are incorporated at frequencies of 66.3% (55/83) and 63.9% (53/83), respectively. The vector-borne BstBI marker 6.6 kb upstream of PACTk and the vector-borne XhoI marker 6.9 kb downstream of PACTk are incorporated at frequencies of 56.6% (47/83) and 42.2% (35/83), respectively. The vector-borne BstBI marker 8.5 kb upstream of PACTk and the vector-borne BamHI marker 9.1 kb downstream of PACTk are incorporated at frequencies of 38.6% (32/83) and 51.8% (43/83), respectively. Finally, the vector-borne AatII marker 11.0 kb upstream of PACTk and the vector-borne BstBI marker 11.2 kb downstream of PACTk are incorporated at frequencies of 21.7% (18/83). These data indicate that even markers as far as 11 kb from the selectable marker can be introduced into the target locus at efficiencies high enough to make their isolation practical. All 13 vector-borne markers were incorporated in four recombinants (D37, D41, D65, and D73), demonstrating that HDAd can simultaneously introduce multiple vector-borne markers as far as 22.2 kb apart (the distance from marker L7 to marker R6 in the chromosome) into the target locus.

Heteroduplex DNA Formation

Examination of the marker patterns in the class II recombinants suggests that extensive hDNA can form during HDAd-mediated gene targeting (Figure 3). For example, in recombinants D77 and D32, hDNA appears to have spanned markers L2 through L7 in the 5′ region of homology, a distance of 9.6 kb. Likewise, in recombinant D29, hDNA appears to have spanned markers R1 through R6 in the 3′ region of homology, a distance of 10.2 kb. The hDNA tracts indicated in Figure 3 for each class II recombinant may underestimate the true length, because MMR may hide evidence of its existence, depending on the direction of repair. For example, class IIB recombinants D67, D41, and D44 showed evidence of hDNA because of their sectored marker. However, there would have been no evidence of hDNA in these recombinants had these sectored markers been repaired (toward either vector or chromosomal sequences) before DNA replication. As another example, class IIA recombinants D17 and D53 showed evidence of hDNA spanning markers R2 through R5, a distance of 5.8 kb. However, if marker R2 had been repaired in the direction of the vector instead of the chromosome, no hDNA would have been evident. That MMR can hide evidence of hDNA can affect interpretation of some results presented in Figure 3. For example, of the 29 class II recombinants, 20 (69%) had evidence of hDNA in the 3′ region of homology, while only 9 (31%) had evidence of hDNA in the 5′ region of homology. This disparity appears to suggest that the 3′ region may be more prone to hDNA formation during HDAd-mediated gene targeting. However, an alternative explanation is that MMR may have hidden evidence of hDNA formation in the 5′ homology in some recombinants, resulting in its underestimation. Another curious finding was that no recombinants had evidence of hDNA in both the 5′ and the 3′ regions of homology. Again, this finding may be explained by MMR hiding evidence of hDNA in one or both homology arms in some recombinants. The total number of class II recombinants may be underestimated due to MMR. A total of 93 markers resided in regions with clear evidence of hDNA in the class II recombinants. Of these, 14 (15%) escaped MMR and were sectored. However, this is likely an overestimate due to the underestimation of hDNA tract lengths because of MMR, as explained earlier. Of the remaining 79 markers within hDNA that were repaired, 38 (48.1%) were repaired toward the vector sequence and 41 (51.9%) were repaired toward the chromosomal sequence, suggesting there was no overall bias in MMR toward the endogenous chromosomal sequence or the vector-borne 2 bp insertion.

DISCUSSION

The abilities to efficiently make multiple targeted changes simultaneously across a large chromosomal region and to correct different

individual mutations widely spread throughout a gene would be powerful and desirable properties of a gene editing vector. Because of their long homology arms, HDAds may offer these important abilities, which we sought to fully assess in this study. To accomplish this, we introduced twelve 2 bp insertions throughout the homology arms of an HDAd so that we could determine how much of the vector-borne sequences are introduced into the recombinant target locus following HR. We found a direct relationship between the distance the genetic marker is from the vector's selectable marker and its frequency of introduction into the target locus. A genetic marker 203 bp from the selectable marker was introduced into all 83 targeted recombinants analyzed. The practical implication of this result is that the selectable marker should be placed as close to the desired genetic modification as possible when designing the vector. However, even markers as far as 11 kb from the selectable marker can be incorporated into the target locus at a relatively high frequency of 21.7%, making their isolation practical. The results also indicated that HDAds are ideally suited for efficiently introducing multiple genetic changes simultaneously over a large genomic region, with 4.8% of the recombinants having all 13 genetic markers introduced into the target locus, the two farthest being 22.2 kb apart. These results support the contention that a single HDAd would be useful to correct many mutations and introduce multiple changes across the large genomic region, as suggested by Liu et al.[4] and Li et al.,[15] respectively.

The HDAd used in this study (HD-23.8-CFTRm-PACTk-DTA) differed from the HDAd used in our prior study (HD-23.8-CFTR-PACTk)[13] in two ways. First, HD-23.8-CFTRm-PACTk-DTA possesses twelve 2 bp insertions, six spread throughout each of the two homology arms, whereas HD-23.8-CFTR-PACTk does not. Second, HD-23.8-CFTRm-PACTk-DTA possesses the poly(A)-less diphtheria toxin A fragment gene to permit passive negative-selection against cells that have randomly integrated the vector.[17] Transduction of 2×10^6 iPSCs at an MOI of 350 vp/cell yielded 177 puro^R colonies, 89.5% of which were correctly targeted with HD-23.8-CFTRm-PACTk-DTA, whereas 144 puro^R colonies, 64.6% of which were correctly targeted, were obtained previously with HD-23.8-CFTR-PACTk.[13] These results indicate that the twelve 2 bp insertions in the region of homology did not adversely affect the efficiency of HDAd-mediated gene targeting and that the poly(A)-less diphtheria toxin A fragment was effective at enriching for targeted recombinants by killing cells harboring random HDAd integrations. These high efficiencies contrast other methods that have been used to correct the ΔF508 mutation in patient-derived stem cells. For example, Crane et al.[16] obtained an average of 29–32 puro^R colonies, 6.3% of which were correctly targeted at the *CFTR* locus following nucleofection of wild-type donor DNA (bearing the same PACTk used in this study) and *CFTR*-specific zinc finger nuclease (ZFN)-expression plasmids into 2×10^6 iPSCs (the same iPSCs used in this study). Schwank et al.[18] obtained 11 to 31 puro^R colonies, of which 9.7% to 27.3% were correctly targeted at the *CFTR* locus following lipofectamine transfection of wild-type donor DNA, Cas9, and *CFTR*-specific guide RNA expression plasmids into 1.4×10^6 intestinal stem cells. Firth et al.[19] analyzed 36 puro^R colonies, of which 16.7% were determined to be targeted at the *CFTR* locus following nucleofection of wild-type donor DNA, Cas9, and *CFTR*-specific guide RNA expression plasmids into 1.4×10^6 iPSCs. However, targeted recombinants in that study were identified by PCR screening of only the 5′ vector-chromosome junction, and this may overestimate the frequency of targeting because we (this study and Palmer et al.[13]) and others[5,16] have shown the existence of aberrantly targeted recombinants that have undergone correct HR in the 5′ region of homology, but not the 3′ region of homology.

Our results revealed that extensive hDNA of up to 10 kb (or more) can form between homologous vector and chromosomal sequences during HDAd-mediated gene targeting and that there does not appear to be a preference for MMR toward the chromosomal sequence or the 2 bp insertion of the vector sequence. However, as described in the Results section, the direction of MMR may hide evidence of hDNA, leading to an underestimation of the number of recombinants with hDNA and the location, as well as the length, of hDNA. Unknown polymorphisms between the CFTR sequences in the vector and those in the chromosome might also inhibit hDNA formation. Our results also show that about 14% of markers within hDNA escape MMR, resulting in a genetically mixed colony. However, this may be an overestimate because of the potential underestimation of hDNA lengths, as explained earlier. Nevertheless, the proportion is high enough to warrant caution, because correctly targeted recombinants may be composed of genetically mixed populations if the region of homology in the vector is not 100% homologous with the chromosome.

The mechanism of gene targeting in yeast and mammalian cells using double-stranded linear donor DNA appears to occur by the ends-out model of HR.[20–23] Because the HDAd genome is linear, the ends-out model might also be the mechanism of gene targeting by HDAd (Figure 5). In this model, DSBs are randomly introduced into the HDAd genome, and those located within the two homology arms could be recombinogenic by initiating strand invasion of homologous chromosomal sequences. This model is consistent with the direct relationship between the frequency of marker introduction and its distance from PACTk shown in Figure 4, because the farther a marker is from PACTk, the greater the probability that it will be lost due to the random location of DSBs. For those recombinants with more than one sectored marker (D55, D68, and D24), this model is also consistent with the observed linkage relationship between these sector markers (all vector-borne markers in one DNA strand and all chromosomal markers on the other DNA strand). In our previous study, we observed that targeting efficiency by HDAd decreased as the length of the homology arms decreased.[13] The ends-out model is consistent with these previous findings, because with longer homology arms, there would be a higher probability that the random DSBs would occur in homologous sequences to generate homologous recombinogenic ends. Conversely, there would be a higher probability that the random DSBs would occur in nonhomologous sequences in vectors with shorter homology arms (HDAd with shorter homology arms must necessarily have longer non-homologous DNA so that the minimal packageable genome size of ~27.7 kb is maintained[24]), leading to lower targeting efficiency and higher incidence of random vector integration, as was observed.[13] This model

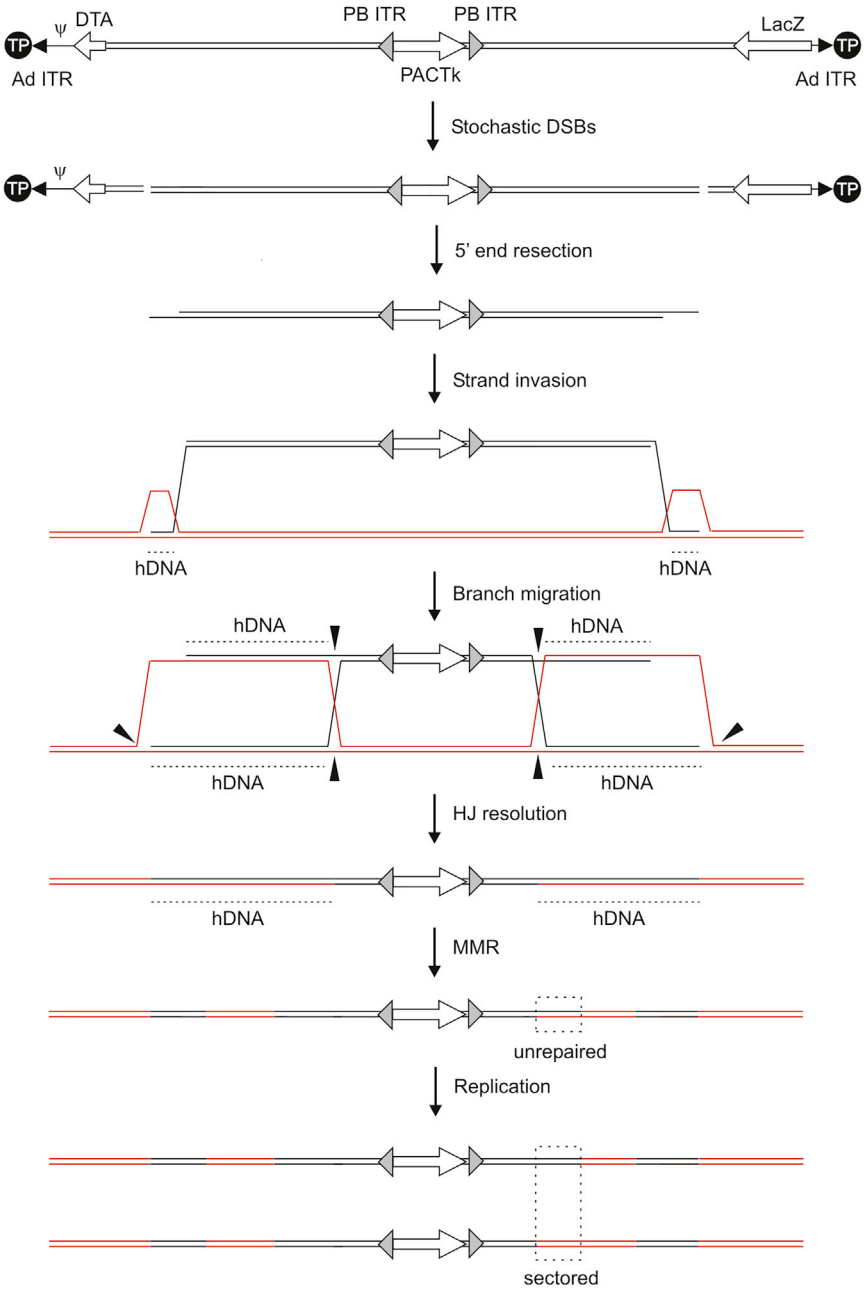

Figure 5. Ends-Out Model for HDAd-Mediated Gene Targeting

Homologous recombination is initiated by random DSBs within the vector's homology arms. Resection of 5′ ends produces recombinogenic single strands with 3′ ends that invade homologous chromosomal sequences to create hDNAs that can be enlarged by branch migration. Resolution of Holliday junctions (HJs) results in HDAd integration. Black triangles indicate cleavage and ligation sites for targeted integration of PACTk. The location of hDNA is indicated by the broken horizontal line. Mismatches in hDNA are repaired toward the vector or chromosomal sequence by mismatch repair (MMR) but occasionally escape repair, leading to sector markers (broken boxes).

can also explain the aberrantly targeted recombinants that were observed in this and our prior study[13] by proposing, for example, that the DSB occurred in homologous sequences in the vector's left arm but that in the right arm, the DSB occurred in nonhomologous sequences (e.g., in LacZ) leading to correct HR in the left arm but nonhomologous end joining (NHEJ) in the right arm. All of the above observations are consistent with the ends-out model, and this provides testable hypotheses important for understanding the mechanism of HDAd-mediated gene targeting so that further improvements in efficiency may be achieved.

In summary, this study clearly demonstrates that HDAds are uniquely capable of efficiently introducing multiple modifications, as far as 22.2 kb apart, simultaneously into a large genomic region. We also found that the closer a vector-borne marker is to the vector's selectable marker, the greater the frequency of its introduction into the recombinant target locus. However, even vector markers as far as 11 kb from the selectable marker can be introduced into the target locus with relatively high efficiency. During this process, extensive hDNA can form between the vector and the chromosome, and while most mismatches are efficiently repaired without bias, a small but signification proportion escapes MMR, leading to genetically mixed colonies.

MATERIALS AND METHODS

HDAds

The HD-23.8-CFTRm-PACTk-DTA was derived from HD-23.8-CFTR-PACTk[13] first by inserting the poly(A)-less diphtheria toxin A fragment gene from pBSDT-AII (Addgene plasmid #27179)[25] into the unique PacI site downstream of the adenoviral packaging signal. Then, the twelve 2 bp insertions shown in Figure 1B were introduced by Q5 site-directed mutagenesis according to the manufacturer's instructions (New England Biolabs). Further cloning details are available upon request. HDAds were produced in 116 cells[22] using the serotype 5 helper virus AdNG163,[27] as described elsewhere.[26,28] HDAd titers were determined by absorbance at 260 nm, as described elsewhere.[24]

Transduction of iPSCs

CF17, the feeder free human cystic fibrosis (CF) iPSC line used in this study, is described elsewhere[16] and was maintained in mTeSR 1 (STEMCELL Technologies) on Matrigel (Corning)-coated plates.

The iPSCs were transduced with HDAd, as described previously.[29] Briefly, 2×10^6 cells were resuspended in 1 mL mTeSR 1 supplemented with Y27632 (Reagents Direct) to 10 μM in a 1.5 mL microfuge tube and transduced with HDAd at an MOI of 350 vp/cell for 1 hr at 37°C with gentle rocking. Following transduction, cells were washed twice with 1 mL mTeSR 1 supplemented with Y27632 to 10 μM and plated into 11 Matrigel-coated wells of 6-well plates in mTeSR 1 supplemented with Y27632 to 10 μM. Puromycin was added to the media to a final concentration of 0.5 μg/mL 48 hr post-transduction, and well-isolated colonies were picked 12 days after transduction. DNA was extracted from colonies for Southern analysis and PCR.

DNA Analysis

DNA extraction and non-radioactive digoxigenin (DIG)-based Southern blot hybridization was performed as described previously.[13] The 5′ region of homology was amplified from the targeted clones using primers 5′-ATGAGGGAAGGACTCATGAGAGGGAAGTAG-3′ and 5′-ATGCTCCAGACTGCCTTGGGAAAAGCG-3′. The latter primer was also used to sequence the 5′ PCR product to determine whether exon 10 bore the wild-type or ΔF508 or ΔI507 sequence. The 3′ region of homology was amplified from the targeted clones using primers 5′-TCTATGGCTTCTGAGGCGGAAAGAACCAG-3′ and 5′-ACGTGTATCTGAGAGTGTTACATGGCCCTG-3′. All PCR amplifications were performed with PrimeSTAR GXL DNA polymerase (Takara/Clontech) with final concentrations of 0.2 mM dinucleotide triphosphate (dNTP) and 0.2 μM of each primer. Thermocycling conditions were as follows; 1 min at 94°C, followed by 30 cycles of 98°C for 10 s and 72°C for 10 s, and a final extension of 10 min at 72°C. For PCR products from recombinants with more than one sectored marker, the linkage relationship between sectored markers was determined by double digestion of the PCR product with the relevant restriction enzymes.

AUTHOR CONTRIBUTIONS

This study was conceived by P.N. This study was designed and implemented by D.J.P., N.C.G., D.L.T., and P.N. This paper was written by P.N.

CONFLICTS OF INTEREST

No competing financial interests exist.

ACKNOWLEDGMENTS

This study was supported by internal Baylor College of Medicine funds.

REFERENCES

1. Ohbayashi, F., Balamotis, M.A., Kishimoto, A., Aizawa, E., Diaz, A., Hasty, P., Graham, F.L., Caskey, C.T., and Mitani, K. (2005). Correction of chromosomal mutation and random integration in embryonic stem cells with helper-dependent adenoviral vectors. Proc. Natl. Acad. Sci. USA *102*, 13628–13633.

2. Suzuki, K., Mitsui, K., Aizawa, E., Hasegawa, K., Kawase, E., Yamagishi, T., Shimizu, Y., Suemori, H., Nakatsuji, N., and Mitani, K. (2008). Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. Proc. Natl. Acad. Sci. USA *105*, 13781–13786.

3. Li, M., Suzuki, K., Qu, J., Saini, P., Dubova, I., Yi, F., Lee, J., Sancho-Martinez, I., Liu, G.H., and Izpisua Belmonte, J.C. (2011). Efficient correction of hemoglobinopathy-causing mutations by homologous recombination in integration-free patient iPSCs. Cell Res. *21*, 1740–1744.

4. Liu, G.H., Suzuki, K., Qu, J., Sancho-Martinez, I., Yi, F., Li, M., Kumar, S., Nivet, E., Kim, J., Soligalla, R.D., et al. (2011). Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. Cell Stem Cell *8*, 688–694.

5. Liu, G.H., Qu, J., Suzuki, K., Nivet, E., Li, M., Montserrat, N., Yi, F., Xu, X., Ruiz, S., Zhang, W., et al. (2012). Progressive degeneration of human neural stem cells caused by pathogenic LRRK2. Nature *491*, 603–607.

6. Aizawa, E., Hirabayashi, Y., Iwanaga, Y., Suzuki, K., Sakurai, K., Shimoji, M., Aiba, K., Wada, T., Tooi, N., Kawase, E., et al. (2012). Efficient and accurate homologous recombination in hESCs and hiPSCs using helper-dependent adenoviral vectors. Mol. Ther. *20*, 424–431.

7. Umeda, K., Suzuki, K., Yamazoe, T., Shiraki, N., Higuchi, Y., Tokieda, K., Kume, K., Mitani, K., and Kume, S. (2013). Albumin gene targeting in human embryonic stem cells and induced pluripotent stem cells with helper-dependent adenoviral vector to monitor hepatic differentiation. Stem Cell Res. (Amst.) *10*, 179–194.

8. Liu, G.H., Suzuki, K., Li, M., Qu, J., Montserrat, N., Tarantino, C., Gu, Y., Yi, F., Xu, X., Zhang, W., et al. (2014). Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. Nat. Commun. *5*, 4330.

9. Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., Goebl, A., Tang, S., Ren, R., Aizawa, E., et al. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. Cell Stem Cell *15*, 31–36.

10. Yoshida, T., Ozawa, Y., Suzuki, K., Yuki, K., Ohyama, M., Akamatsu, W., Matsuzaki, Y., Shimmura, S., Mitani, K., Tsubota, K., and Okano, H. (2014). The use of induced pluripotent stem cells to reveal pathogenic gene mutations and explore treatments for retinitis pigmentosa. Mol. Brain *7*, 45.

11. Zhang, W., Li, J., Suzuki, K., Qu, J., Wang, P., Zhou, J., Liu, X., Ren, R., Xu, X., Ocampo, A., et al. (2015). Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. Science *348*, 1160–1163.

12. Yamamoto, H., Ishimura, M., Ochiai, M., Takada, H., Kusuhara, K., Nakatsu, Y., Tsuzuki, T., Mitani, K., and Hara, T. (2016). BTK gene targeting by homologous recombination using a helper-dependent adenovirus/adeno-associated virus hybrid vector. Gene Ther. *23*, 205–213.

13. Palmer, D.J., Grove, N.C., Ing, J., Crane, A.M., Venken, K., Davis, B.R., and Ng, P. (2016). Homology requirements for efficient, footprintless gene editing at the *CFTR* locus in human iPSCs with helper-dependent adenoviral vectors. Mol. Ther. Nucleic Acids *5*, e372.

14. Holkers, M., Maggio, I., Henriques, S.F., Janssen, J.M., Cathomen, T., and Gonçalves, M.A. (2014). Adenoviral vector DNA for accurate genome editing with engineered nucleases. Nat. Methods *11*, 1051–1057.

15. Li, M., Suzuki, K., Kim, N.Y., Liu, G.H., and Izpisua Belmonte, J.C. (2014). A cut above the rest: targeted genome editing technologies in human pluripotent stem cells. J. Biol. Chem. *289*, 4594–4599.

16. Crane, A.M., Kramer, P., Bui, J.H., Chung, W.J., Li, X.S., Gonzalez-Garay, M.L., Hawkins, F., Liao, W., Mora, D., Choi, S., et al. (2015). Targeted correction and restored function of the CFTR gene in cystic fibrosis induced pluripotent stem cells. Stem Cell Reports *4*, 569–577.

17. Yagi, T., Ikawa, Y., Yoshida, K., Shigetani, Y., Takeda, N., Mabuchi, I., Yamamoto, T., and Aizawa, S. (1990). Homologous recombination at c-fyn locus of mouse embryonic stem cells with use of diphtheria toxin A-fragment gene in negative selection. Proc. Natl. Acad. Sci. USA *87*, 9918–9922.

18. Schwank, G., Koo, B.K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K., et al. (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell Stem Cell *13*, 653–658.

19. Firth, A.L., Menon, T., Parker, G.S., Qualls, S.J., Lewis, B.M., Ke, E., Dargitz, C.T., Wright, R., Khanna, A., Gage, F.H., and Verma, I.M. (2015). Functional gene correction for cystic fibrosis in lung epithelial cells generated from patient iPSCs. Cell Rep. *12*, 1385–1390.

20. Hastings, P.J., McGill, C., Shafer, B., and Strathern, J.N. (1993). Ends-in vs. ends-out recombination in yeast. Genetics *135*, 973–980.

21. Li, J., Read, L.R., and Baker, M.D. (2001). The mechanism of mammalian gene replacement is consistent with the formation of long regions of heteroduplex DNA associated with two crossing-over events. Mol. Cell. Biol. *21*, 501–510.

22. Langston, L.D., and Symington, L.S. (2004). Gene targeting in yeast is initiated by two independent strand invasions. Proc. Natl. Acad. Sci. USA *101*, 15392–15397.

23. Kan, Y., Ruis, B., Lin, S., and Hendrickson, E.A. (2014). The mechanism of gene targeting in human somatic cells. PLoS Genet. *10*, e1004251.

24. Parks, R.J., and Graham, F.L. (1997). A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. J. Virol. *71*, 3293–3298.

25. Aoyama, M., Agari, K., Sun-Wada, G.H., Futai, M., and Wada, Y. (2005). Simple and straightforward construction of a mouse gene targeting vector using in vitro transposition reactions. Nucleic Acids Res. *33*, e52.

26. Palmer, D., and Ng, P. (2003). Improved system for helper-dependent adenoviral vector production. Mol. Ther. *8*, 846–852.

27. Palmer, D.J., and Ng, P. (2004). Physical and infectious titers of helper-dependent adenoviral vectors: a method of direct comparison to the adenovirus reference material. Mol. Ther. *10*, 792–798.

28. Palmer, D.J., and Ng, P. (2008). Methods for the production of helper-dependent adenoviral vectors. Methods Mol. Biol. *433*, 33–53.

29. Palmer, D.J., Grove, N.C., and Ng, P. (2016). Helper virus-mediated downregulation of transgene expression permits production of recalcitrant helper-dependent adenoviral vector. Mol. Ther. Methods Clin. Dev. *3*, 16039.