REVIEWS

PRECISE HIT: ADENO-ASSOCIATED VIRUS IN GENE TARGETING

Ana Vasileva and Rolf Jessberger‡*

Abstract | Vectors based on the adeno-associated virus (AAV) have attracted much attention as potent gene-delivery vehicles, mainly because of the persistence of this non-pathogenic virus in the host cell and its sustainable therapeutic gene expression. However, virus infection can be accompanied by potentially mutagenic random vector integration into the genome. A novel approach to AAV-mediated gene therapy based on gene targeting through homologous recombination allows efficient, high-fidelity, non-mutagenic gene repair in a host cell.

VECTORS
Nucleic-acid sequences derived from bacterial plasmids or viruses, used for delivery of exogenous DNA or RNA into cells.

TRANSGENE
Artificially reconstituted gene, which can be expressed either pre- or post-integration into the genome; often encodes a therapeutic or a marker gene.

SILENCING
In this context, complete or partial shut-off of transgene expression owing to factors acting at the locus of vector integration and/or the structure of the locus.

*Department of Gene and Cell Medicine, Mount Sinai School of Medicine, New York, USA.
‡Institute for Physiological Chemistry, Dresden University of Technology, Dresden, Germany.
Correspondence to R.J.
e-mail: Rolf.Jessberger@ mailbox.tu-dresden.de*

Contemporary gene-therapy strategies and gene-therapy trials mostly use VECTORS that integrate randomly or quasi-randomly into the host cells' genome (reviewed in REFS 1–3). Although this is an efficient way to stably deliver genes into the target cell, it has several crucial disadvantages (TABLE 1). One of these disadvantages is the variable site and frequency of integration of the TRANSGENE — multiple copies can integrate into the host genome and mutagenize it, with potentially devastating consequences for the cell and for the organism. Cell death is probably the most harmless consequence, whereas cell transformation is a worrisome outcome. Human gene-therapy trials using retroviral vectors indicate that quasi-random integration of transgenes can indeed induce malignant transformation in transduced cells[4].

Another disadvantage of the random integration of a transgene is the unpredictability of its stability and its expression. The genomic locus in which the vector integrates can have profound effects on the level of transgene expression, as it can completely SILENCE the transgene, or it can increase or decrease its expression. These effects cannot be easily avoided even by sophisticated vector design. Even the inclusion of the gene's own promoter and/or enhancer region in the transgenic vector construct does not necessarily prevent these host-genome-mediated effects, as the surrounding chromatin can override the activity of the original regulatory regions.

Gene targeting by HOMOLOGOUS RECOMBINATION (HR), however, lacks many of these shortcomings. In this process, the transgene recombines with its natural locus in the host genome, thereby ensuring correct transcription. Also, after HR, the targeted modification of the chromosomal locus is stable, whereas randomly integrated sequences might be lost over time. Gene correction by targeting a sufficient number of cells containing one functional gene copy can eliminate dominant mutations causing genetic disease, which would be difficult to overcome by the competitive but variable expression of randomly inserted wild-type genes. Alternatively, the mutant copy can be knocked out to ensure expression of the existing functional allele. Furthermore, the correction or targeted disruption of a dominant mutation can be heritable if stem cells or germline cells are targeted.

For successful gene targeting, the vector must find the correct chromosomal target, which is a major task considering that the human genome comprises approximately three billion base pairs (bp). The only known way for the transgene to react with its specific host genomic target is by DNA recombination between homologous sequences that are present on the vector and the target gene. However, compared with bacteria and yeast, targeted HR in mammalian cells is an inefficient process, with frequencies of 1 in 10^6 or lower[5,6]. In a mammalian cell, instead of undergoing the complex task of finding the homologous chromosomal sequence, vector DNA

often enters into other pathways of DNA repair, notably the end-joining pathway called NON-HOMOLOGOUS END JOINING (NHEJ). Here, the vector DNA recombines with unrelated sequences and can integrate almost anywhere in the genome. The frequency of random integration events is about a thousand-fold higher than that of HR. NHEJ is mediated by proteins that are different and mostly independent from those needed to carry out HR[7]. Many of the important components of the NHEJ and the HR machineries have been identified (BOX 1), but a clear understanding of how the balance between homologous and non-homologous DNA recombination is regulated has not yet been achieved.

The design of vectors for homologous gene targeting is demanding. Standard vectors contain two sequences that are perfectly homologous and ISOGENIC to the genomic target sequence. These stretches are typically called the long and the short arm and require several thousand bp of genomic DNA. The two arms often flank a selectable marker gene such as the neomycin resistance gene (neo), which interrupts the target gene in the host genome to produce a 'knockout'[8]. There are many variations of this scheme, and not only knockouts but also knock-ins, or gene replacements, can be generated, in which a defective host gene can be replaced by its intact copy. There have been many attempts to improve the low targeting efficiency of these reactions, including microinjection and other DNA-delivery methods[9,10], co-transfection with recombinase genes[11–13], pre-engineering of induced double-strand breaks (DSBs) at the target sequence[14–16], delivery by adenovirus[17,18] and other methods (see Supplementary information S1 (table)). However, these strategies have not greatly increased the frequency of repair or are not easily applicable to gene therapy.

In 1998, a surprising observation was reported: DNA vectors that were based on the adeno-associated virus (AAV) could target homologous chromosomal DNA sequences in mammalian cells at astonishingly high frequencies[19]. In up to 1% of cells TRANSDUCED with such vectors, precise, site-specific modifications were introduced. This frequency is several orders of magnitude higher than those achieved with other gene-targeting strategies[5]. This review discusses the important features and specific characteristics of the recombinant AAV (rAAV) system that are responsible for its impressive gene-targeting frequencies, and proposes a possible mechanism of targeting by rAAV.

The AAV virus

AAV is a single-stranded DNA animal virus that belongs to the *Dependovirus* genus of the *Parvoviridae* family. Nine subtypes of the virus (AAV1–9) have been described for which humans are the primary host. Despite seropositivity of 80% in the human population for the most common subtype, AAV2, no known pathology is associated with this virus. AAV is a small icosahedral virus with a single copy of a 4.7-kb DNA genome, with characteristic termini consisting of palindromic repeats that fold into a hairpin shape (inverted terminal repeats (ITRs); FIG. 1a). The wild-type genome contains two open reading frames (ORFs) termed *rep* and *cap*, which encode non-structural and structural proteins, respectively. In the absence of a helper virus such as adenovirus or herpes simplex virus, AAV integrates into the human genome at a specific region on chromosome 19, designated AAVS1, and persists in a latent form. Following infection with helper virus that provides factors necessary for active replication, the virus can enter a lytic cycle.

Recombinant AAV as a gene-delivery vector

The defective replication and non-pathogenic nature of wild-type AAV triggered the rapid development of rAAV derived from AAV2. In these vectors, the *rep* and *cap* ORFs are replaced with a gene-expression cassette of interest, whereas for vector production, the *rep* and *cap* gene products as well as helper-virus elements are supplied in *trans* (FIG. 1b). rAAV can transduce various cell types in different species, and several studies have reported the successful application of rAAV vectors in mouse, rat, monkey and other animal models of human disease[20]. Moreover, rAAV can transduce cells both *in vivo* and *ex vivo*.

When rAAV is used as a gene-delivery vehicle, most transgene expression results from extrachromosomal viral genomes that persist as double-stranded circular or linear EPISOMES[21,22]. This might be useful in gene-therapy settings in which post-mitotic quiescent tissues are pursued, but sufficient episomal expression levels would not be sustained in rapidly dividing cells such as haematopoietic cells or stem cells and, indeed, in one mouse study, approximately 90% of the rAAV episomes were lost soon after hepatocyte proliferation induced by partial hepatectomy[23].

Stable transgene expression can be ensured through the integration of viral episomes by NHEJ.

HOMOLOGOUS RECOMBINATION
(HR). Substitution (reciprocal exchange or replacement) of a segment of DNA by another that is identical (homologous). Occurs naturally during meiotic recombination and DNA repair.

NON-HOMOLOGOUS END JOINING
(NHEJ). Process of joining two broken DNA ends on the basis of limited or no homology between two DNA molecules, which can result in incorrect recovery of the sequence of a broken DNA molecule. Also functions to create genetic diversity during immune development.

ISOGENIC
Genetically identical (except for sex).

TRANSDUCTION
In this context, equivalent to infection with a virus used as a gene-therapy vector.

EPISOME
Non-integrated, circular, double-stranded DNA molecule with the capacity to integrate into the chromosome of the host cell.

Table 1 | **Advantages and disadvantages of gene addition and gene targeting approaches**

Criteria/characteristics	Gene targeting (homologous recombination)	Random integration (non-homologous recombination)
Sites of integration	Predictable	Unpredictable
Mutagenicity	High fidelity	Potentially mutagenic
Number of events	Maximum two	Multiple
Regulation of expression	Endogenous	Variable, exogenous
Vector design	Complex (laborious)	Simple
Efficiency	Very low	1000-fold higher than homologous recombination

Box 1 | **Major pathways of DNA double-strand break repair**

Cells preserve their genome integrity by repair of double-strand breaks (DSBs), which can be caused by factors such as stalled DNA replication or by environmental mutagens such as ionizing radiation (IR). The two main pathways of DSB repair in eukaryotes, non-homologous end joining (NHEJ) and homologous recombination (HR), were first studied in radiosensitive *Saccharomyces cerevisiae* mutants[86,87] and are well conserved in mammals.

In mammalian cells, DSBs induced during the G1 phase of the cell cycle are repaired predominantly by the NHEJ pathway (see figure, part **a**). Here, the heterodimer KU70/KU86, a component of the DNA-dependent protein kinase (DNA-PK) holoenzyme, binds DNA ends and recruits the catalytic subunit of DNA-PK (DNA-PK$_{cs}$). A conformational change in DNA-PK$_{cs}$ brings the globular domains together, and the two complexes form a synapse in which the DNA ends reside and are protected from modifying enzymes. Autophosphorylation of DNA-PK exposes the ends to such enzymes, and its kinase activity is required to activate the nuclease Artemis, which processes the ends. Last, the XRCC4–DNA ligase IV complex interacts with the DNA-PK holoenzyme and completes the repair process by ligating the ends.

Single-strand annealing (SSA) and canonical HR are initiated when activation of ATM kinase signalling downstream of DSBs recruits the MRE11–RAD50–NBS1 (MRN) complex to DNA ends. Processing of the ends by a 5′–3′ exonuclease leaves free 3′ overhangs, which form secondary structures.

SSA (see figure, part **b**) is restricted to DSBs that occur between repeats. After resection of the broken ends, the RAD52 heptamer brings together complementary single-stranded regions. The unpaired single-stranded DNA (ssDNA) overhangs are removed by the structure-specific endonuclease ERCC1/XPF and the gaps are filled by DNA synthesis and ligated. The deletion of a repeat and flanking DNA is potentially mutagenic.

In the HR pathway (see figure, part **c**), secondary DNA structures are removed by the ssDNA-binding protein replication protein A (RPA), which interacts directly with RAD52 and is displaced by BRCA2. Through its BRC repeats, BRCA2 brings along the filament forming protein RAD51, which acquires high-affinity binding for ssDNA in the presence of RAD52. RAD51 paralogue complexes such as the XRCC3/RAD51C form foci at DSBs prior to RAD51 and are thought to help to recruit RAD51 molecules. Strand invasion and homologous pairing of the RAD51–DNA filament is assisted by unwinding of the homologous region by the RAD54 helicase and other proteins. The sequence on the broken DNA is recovered by either repair synthesis using the intact DNA as a donor (gene conversion, non-crossover) or by DNA exchange (crossing over) and resolution by resolvases following formation of a Holliday junction.
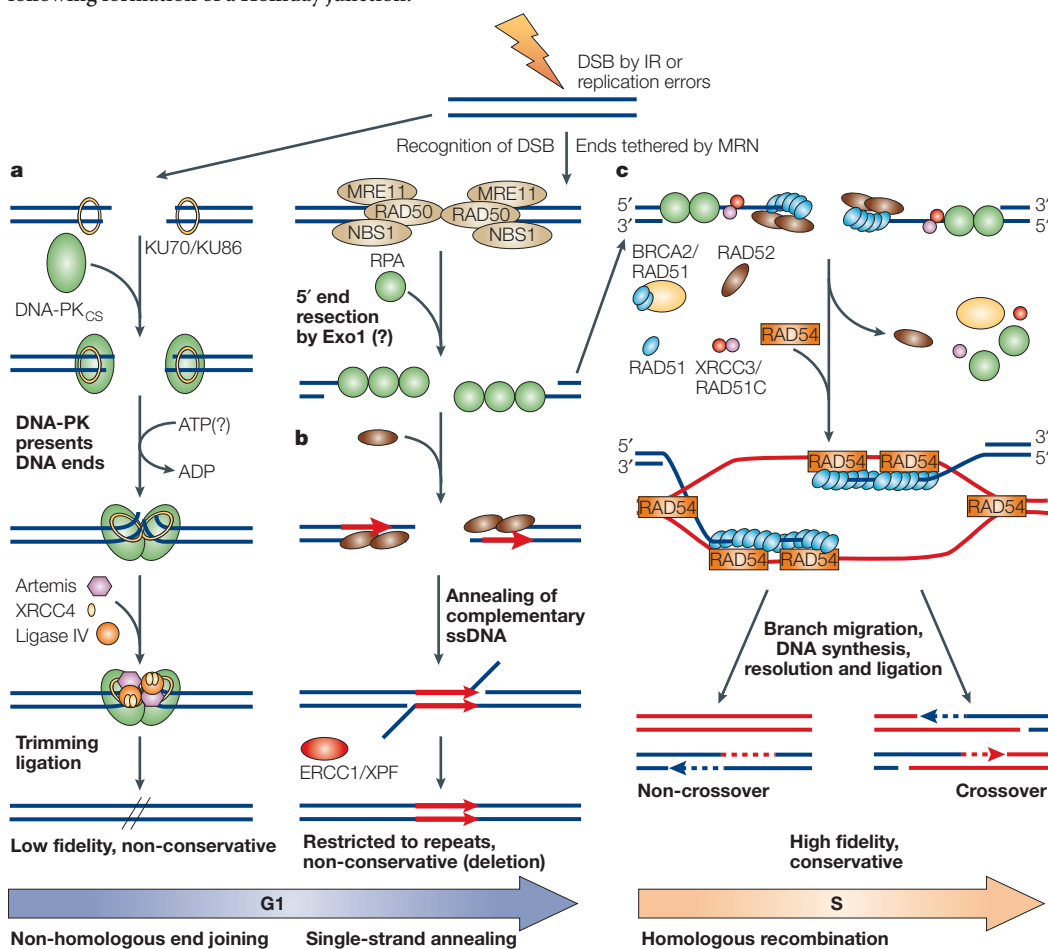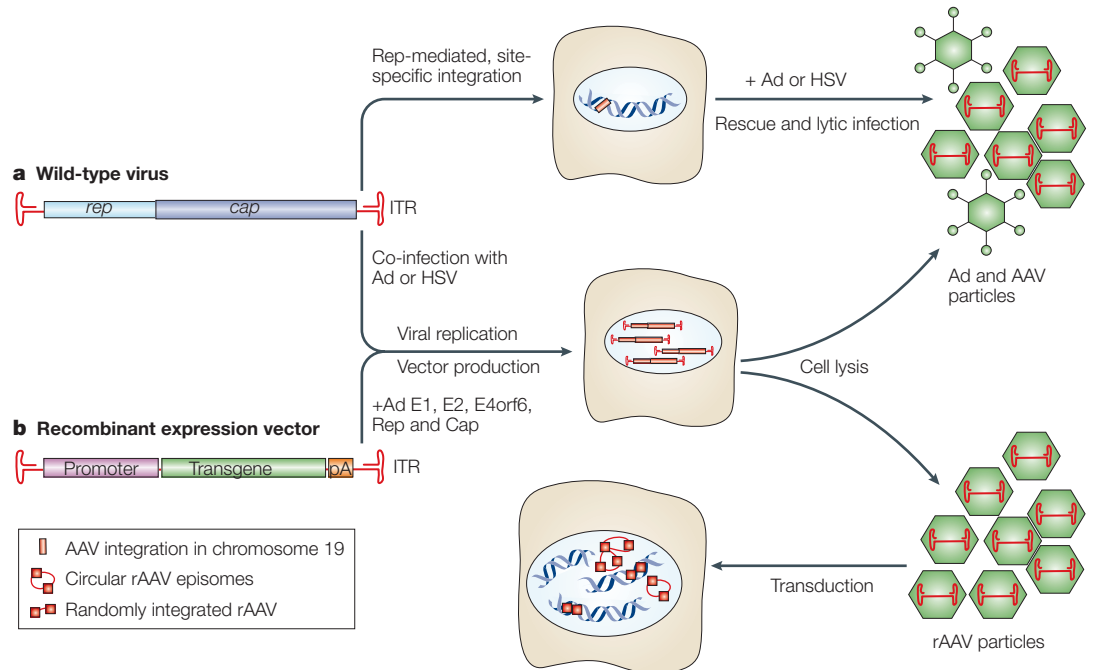
Figure 1 | **Biology of wild-type and recombinant adeno-associated virus. a** | The structure of the wild-type adeno-associated virus (AAV) is shown. A single-stranded DNA genome is encompassed by palindromic inverted terminal repeats (ITRs). The *rep* open reading frame (ORF) encodes proteins that are involved in viral replication, and the *cap* ORF encodes proteins that are necessary for viral packaging. AAV integrates into the human genome at a specific locus on chromosome 19 (red) and persists in a latent form. It can exit this stage only if the cell is co- or super-infected with helper virus such as adenovirus (Ad) or herpes simplex virus (HSV), which provide factors necessary for active AAV replication. **b** | The generic gene-delivery vector based on AAV is depicted. The viral genome is replaced by an expression cassette, which usually consists of a promoter, transgene and polyA (pA) tail. For production of the recombinant virus (rAAV), Rep and Cap proteins as well as Ad or HSV elements (Ad E1, E2 and E4orf6) have to be provided in *trans*. Examples of intracellular forms of the delivery vector that are responsible for transgene expression following transduction with rAAV (double-stranded circular episomes and randomly integrated vector genomes) are depicted in red.

However, data from transduced mouse liver indicate that this is potentially mutagenic. Up to 72% of rAAV insertions occur preferentially in actively transcribed genes and are associated with genomic deletions[24]. Although a locus that contains rRNA repeats is an important 'hot spot' for virus integration, half of the integration events occurred near transcription start sites or CpG islands, which can potentially perturb the regulation of flanking genes[25]. Also, about 3.5% of integrations occurred in potentially oncogenic genes. Integration sites in human HeLa cells also clustered in gene-rich regions[26].

The ability of wild-type AAV, or of rAAV, to integrate site-specifically in chromosome 19 in the presence of the Rep protein was considered potentially useful for gene therapy[27–29] — in principle, it should be possible to direct all integration events to a known, perhaps harmless site. However, the overexpression of wild-type Rep upregulates the expression of many cellular and viral genes and induces apoptosis[30], which limits therapeutic applications of chromosome-directed AAV vectors. Also, it has been shown that the genomic locus into which Rep-mediated integration occurs codes for a cellular protein that is ubiquitously expressed in human adult tissues, with particularly high expression in cardiac tissue[31]. The precise function of this protein, myosin-binding subunit 85 (MBS85), is so far

unknown. Only latently infected cells, in which one allele is disrupted, were observed, perhaps indicating that complete loss of MBS85 is detrimental. It seems as if, for practical purposes, site-specific, Rep-mediated chromosome-directed integration by AAV has at least to be further characterized.

The small packaging capacity of the traditional rAAV vectors also presents a limitation to its use. Theoretically, rAAV vectors can be used only if the length of a transgene including its regulatory elements does not exceed 5 kb[32]. For many diseases caused by larger genes, this is a significant restriction and, therefore, efforts to reconstruct larger genes *in vivo* have expanded the system's usefulness. Single-stranded rAAV genomes form circular intermediates including concatemers that can undergo intermolecular recombination[33], and the simultaneous administration of two vectors, each carrying one half of a gene, leads to concatemerization and allows expression of the entire gene by TRANS-SPLICING[34]. Several groups have tried variants of this approach[35–38], but these and other methods need further development. Fortunately, for efficient gene repair, gene targeting by rAAV uses stretches of homology that are smaller than the packaging limit of the virion (see below).

These data illustrate the ongoing problems for the clinical use of rAAV and render alternative approaches such as gene targeting more attractive.

TRANS-SPLICING
A reconstruction of a larger transcript by joining the 5′ and 3′ portions of a transcript from two independent viruses (co-adminstered to the same tissue) that have formed intermolecular concatemers.
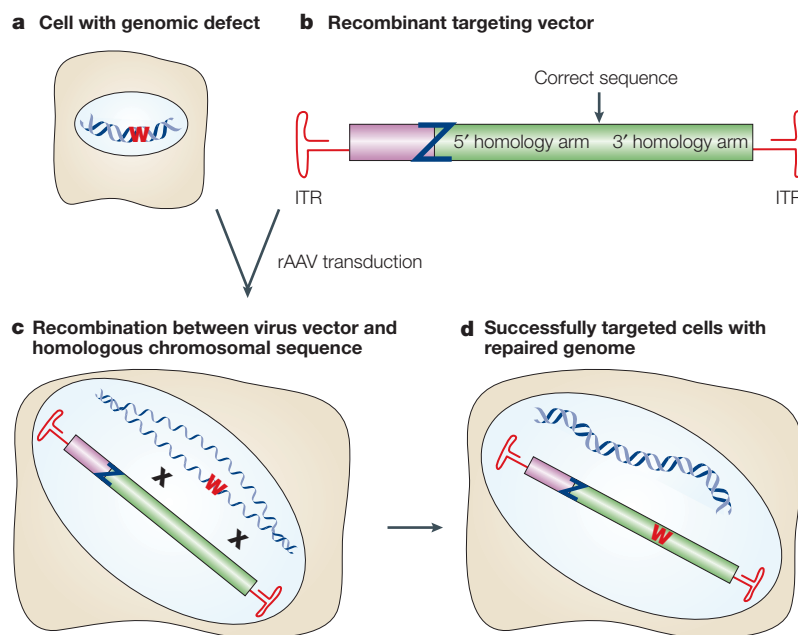
a Cell with genomic defect

b Recombinant targeting vector

Correct sequence

5′ homology arm   3′ homology arm

ITR

ITR

rAAV transduction

c Recombination between virus vector and homologous chromosomal sequence

d Successfully targeted cells with repaired genome

Figure 2 | **Gene targeting with rAAV vectors.** A non-functional mutant of a reporter gene (W in the figure) is introduced or is present in a cell (**a**). In the recombinant targeting vector (**b**), the viral genes are replaced by a sequence that is homologous to the chromosomal locus targeted for modification. Usually, the modification is introduced between stretches of homology called 5′ and 3′ homology arms. Z designates an inactivating mutation in the viral repair template. This recombinant adeno-associated virus (rAAV) is used to transduce the cells. Recombination with the chromosomal target (**c**) can result in repair of the defect and recovery of a healthy cell (**d**). The frequency of gene targeting is determined as the fraction of infected cells that expresses a functional reporter. Stable integration is confirmed by antibiotic selection and Southern analysis. Crosses (X) mark regions of homology between the chromosomal and the viral DNA. ITR, inverted terminal repeat.

rAAV as a gene-targeting vector

To improve the frequencies of gene targeting obtained with other systems and to take advantage of the high transduction potential of rAAV vectors, Russell and colleagues tested the hypothesis that rAAV vectors could recombine with homologous sequences present on a host chromosome (FIG. 2) to introduce defined modifications at high frequencies[19].

Successful gene targeting by rAAV was initially reported in two independent systems. One consisted of the targeted correction of a null mutation in an integrated exogenous *neo* gene that was used as a targeting substrate in HeLa cells (FIG. 2). The transduced rAAV vector carried another mutant *neo* gene with an inactivating insertion of 3 bp. HR between the two mutant genes restored the functional gene, giving rise to G418-resistant colonies. The total length of homology between the substrate and the targeting vector was 2.7 kb, which is about four times less than the length required for efficient targeting with conventional systems. Recombination frequencies were as high as 0.1%, about 2 logs higher than frequencies previously observed with adenoviral and retroviral vectors[39], and 3 to 4 logs higher than frequencies with the identical vector introduced by transfection as a plasmid.

MULTIPLICITY OF INFECTION
Synonym of virus load, the number of viral particles or infectious units per infected cell. Usually represented by the number of viral genomes — gcp/cell (genome-containing particles per cell) or TU/cell (transducing units per cell).

The second system was used to show the targeted disruption of an endogenous human locus. Cells that lack hypoxanthine phosphoribosyltransferase (HPRT) can be grown in the presence of 6-thioguanine (6-TG), and therefore, single-event targeting of this locus on the X chromosome can be studied in XY male cells. Transduction with rAAV vectors introduced a 4-bp insertion into the *HPRT* gene in either normal male fibroblasts or HT1080 human fibrosarcoma cells that had a pseudodiploid male karyotype.

In both systems, targeting rates depended on cell type and MULTIPLICITY OF INFECTION (MOI) and ranged between 0.02% and 0.42%, rarely even up to 1%. High MOI resulted in the highest frequencies of recombination, which indicates that increased DNA uptake during transduction and a high vector-to-target ratio might improve the chances of finding the target or might more efficiently induce a DNA-repair response. Infection with the highest MOI resulted in 10% random integration in addition to 1% HR. Although this percentage of random integration is significant, the ratio of random integration to targeting is 2–3-fold lower than that found with other transduction systems such as retroviruses[39].

These initial results defined a broader scope for rAAV applications and encouraged further studies to establish the conditions that affect rAAV-mediated gene targeting in human cells (BOX 2).

The rAAV vector and host-specific transduction

Similar to other rAAV vectors, rAAV targeting vectors have a broad host range and can deliver DNA to many primary cell types. However, comparisons of results from animal models with those from primary human cells, which are the natural host of the virus, are rare, and therefore it remains unclear whether the results from animal studies will be predictive for responses in human cells. A recent study compared rAAV-mediated correction of a mutant enhanced-green-fluorescent-protein (*eGFP*) gene in both a human embryonic kidney HEK293 cell line and primary mouse embryonic fibroblasts (PMEFs). Targeting efficiency in HEK293 cells was 0.1%, whereas efficiency in PMEFs was significantly lower (~0.006%)[40]. The decrease in recombination frequency was attributed to the 20–30-fold decrease in rAAV transduction levels in PMEFs compared with HEK293 cells. Similarly, the expression of luciferase from a control AAV vector was 10-fold higher in normal human fibroblasts than in PMEFs. Previous findings indicate that the human, but not the murine, fibroblast growth-factor receptor is the co-receptor for AAV type 2 infection, and that intracellular trafficking of the virus in murine fibroblasts is impaired[41,42]. Gene-targeting studies carried out using either AAV or another member of the *Parvoviridae*, minute virus of mice (MVM), confirmed the importance of host-specific transduction, as in human cells, rAAV-mediated targeting was 5-fold higher than the frequency observed for MVM vectors[43].

In most systems, an exogenously inactivated, that is, mutated, marker gene is stably integrated into the genome of a cell line to define the basics of gene repair by rAAV targeting vectors. Genes such as *neo*, alkaline phosphatase (*AP*) and enhanced green fluorescent protein (*eGFP*) can be conveniently detected and allow the comparison of targeting efficiencies obtained with different virus vectors, as well as the optimization of parameters relevant to gene targeting.

These studies can be differentiated into selective and non-selective systems based on the target gene (see Supplementary information S2 (table)). Evaluation of the frequencies of *neo* gene correction by the number of cells surviving on G418 or the frequency of hypoxanthine guanine phosphoribosyl transferase (*HPRT*) disruption by survival on 6-thioguanine (6-TG) involves lengthy selection of the transduced cells, which renders the precise accounting of the proportion of surviving cells difficult. Therefore, systems have been developed for the correction of mutant *eGFP* or *AP* that offer detection of functional gene expression under non-selective conditions in living or histochemically stained cells. These systems allow more accurate estimation of the number of correction events in the population. Indeed, random integration frequencies were 10-fold lower when antibiotic resistance rather than eGFP expression was used as a read-out[64].

An inherent disadvantage of randomly integrated transgene targets is that their expression is controlled by regulatory elements at their random chromosomal integration sites. The use of endogenous genes in their natural chromosomal context such as *HPRT*, the TYPE I COLLAGEN locus, *COL1A1* and the prion protein gene (*PRNP*) avoids this problem, but in most cases these systems are limited to the study of gene disruption and are not appropriate for the study of mutational repair[57,82,85]. Nevertheless, the success of these approaches shows that rAAV targeting might be amenable for clinical use.

The wide variation of recombination frequencies (0.006–1%) obtained using these different approaches indicates that rAAV gene-targeting efficiency does not depend primarily on the individual target, but instead reflects the specific parameters of the experimental system such as the extent of homology, cell type, chromosomal context and virus load.

The single-stranded DNA genome and targeting

Transducing AAV vector DNA instead of transfecting a plasmid-type DNA containing the same sequences yielded a 1,000 times higher gene-targeting frequency. Therefore, the way the cell takes up the DNA affects the targeting frequency. The efficiency of DNA uptake alone, however, is not the decisive factor in targeting frequency, as it has been shown that although more than 5% of plasmid-transfected cells can be positive for the vector DNA, this does not always result in a high incidence of gene targeting[19]. Therefore, intracellular mechanisms that recognize features of rAAV DNA and not the specific rAAV DNA sequence are crucial for the efficiency of gene targeting by rAAV.

One can speculate that the unique structure of the single-stranded AAV genome, with its T-shaped hairpin ends, has a central role in the targeting process. AAV genomes smaller than half the wild-type genome can be packaged either as wild-type-like single-stranded monomers or dsDNA dimers known as self-complementary AAV (scAAV). Transduction using monomers alone or a combination of monomer and scAAV vectors showed a correlation between targeting rates and the amount of monomer vector genomes. By contrast, no significant contribution could be attributed to the presence of dimer genomes[44]. These data indicate that the linear single-stranded monomers mimic DNA damage and induce DNA-repair pathways. The strand preference

for the targeting substrate reported for MVM targeting vectors that have asymmetric ITRs and that package only positive or negative strands[43] provides further evidence, albeit from a different system, for the hypothesis that the presence of free DNA ends at the ITRs and the single-stranded viral genome are required for efficient rAAV-mediated targeting.

Genotoxic stress, the cell cycle and targeting

rAAV vectors preferentially transduce actively proliferating primary human fibroblasts[45], and high targeting frequencies for rAAV have been reported mostly in transformed or rapidly proliferating cells such as HeLa, HEK293, HT1080 and human primary fibroblasts[19]. By contrast, no targeting was observed in mouse tibialis muscle, which is composed mostly of quiescent cells[40]. Therefore, in addition to recombination and/or repair mechanisms, active replication of the cellular genome might be required for efficient targeting. Transduction can be markedly enhanced by the pre-treatment of target cells with physical and chemical agents that perturb DNA metabolism and induce DNA repair[46–48]. Genotoxic stress induced by ETOPOSIDE or inhibition of DNA synthesis by HYDROXYUREA (HU) — both of which arrest cells in S phase of the cell cycle — are known to increase random integration[48], whereas *HPRT* targeting rates have been shown to decrease after HU treatment[49]. Although this indicates that the DNA-recombination pathways required for targeting might be distinct from those supporting random integration, the effect of inhibitors of DNA synthesis on gene-targeting rates is unclear, as in other studies with these agents, the efficiency of repair of alkaline phosphatase (*AP*) and *eGFP* targets was not altered[40,44].

AP staining or 6-TG resistance indicative of correct gene repair has been correlated with bromodeoxy-uridine (BrdU) incorporation, indicating that cells that have replicated their genomes have undergone targeting[49]. In another study, the enrichment of cells in S phase by a double thymidine block applied prior and during transduction of rAAV resulted in 2-fold higher targeting rates as assessed 7 days post-infection compared with untreated, non-synchronized control populations[40]. This increase was accompanied by greater than 3-fold higher transduction rates, whereas cells blocked in other phases of the cell cycle had about half the targeting rates and a ~3-fold reduction in transduction efficiency. This increase in rAAV targeting rates observed after thymidine block might, however, be independent of the cell cycle and could represent enhanced DNA-repair rates in the arrested population. It is important to note that none of these experiments precisely define the phase of the cell cycle in which the actual recombination reaction takes place, and the lag between cell-cycle arrest, infection and the analysis is of concern. Nevertheless, these studies indicate that therapeutic targeting would be most efficient in actively cycling cells, and that the mechanism of rAAV targeting involves DNA repair.

It is important to note that DSBs in cells synchronized in S phase are repaired by HR with a frequency 39 times

higher than in those arrested in M phase, and 24 times higher than in cells arrested in G_1/G_0 phase[50], and key proteins of the HR repair pathway are not expressed in quiescent G_0 cells[51,52]. Also, cells deficient in the NHEJ pathway are most sensitive to ionizing radiation during the G1 phase of the cell cycle[53]. Based on this rather simple correlation, it seems that HR (BOX 1) is the most likely DSB repair pathway responsible for rAAV gene targeting.

Types of modification introduced by rAAV

The potential of rAAV-mediated targeting for therapeutic applications as well as for basic research depends on the types of modification that can be introduced into the targeted genome, the fidelity of the process and the accessibility of different chromosomal locations. A valid comparison of the frequency of repair of various mutated loci can only be ensured if parameters such as viral MOI, cell line, virus exposure time and other experimental conditions are kept constant. Two reports have employed such strategies and compared the frequency of rAAV-mediated repair of deletions, insertions or base substitutions in a *neo* or an *AP* gene that was introduced at various chromosomal positions in a polyclonal cell population by retroviral vector transduction[44,54]. However, as retrovirus integration of the marker gene is not unbiased but favours transcriptionally active regions[55] (that is, open chromatin configurations), certain position effects in these experiments cannot be excluded.

The rAAV targeting vectors corrected a 1-bp insertion in the *neo* gene at about 10 times lower frequency than deletions located at the same site[54]. The authors suggest that mismatch repair enzymes similar to the enzymes that repair single-bp substitutions might recognize specific mutations during heteroduplex formation between the vector and the target DNA[56]. However, more experiments will be required to define the mechanism of single base correction and to validate this hypothesis.

Several other reports have compared the correction frequencies of various mutations such as deletions and insertions that were introduced into either an *AP* or an *HPRT* gene[44,56,57]. These studies confirmed that deletions in the target are repaired at a higher rate than insertions, especially if the rAAV vector and substrate share regions of short homology (see Supplementary information S3 (table)). When cells were rendered 6-TG resistant by the targeted insertional mutagenesis of *HPRT*, the targeting frequencies for large insertions were higher than for smaller insertions. Furthermore, these studies confirmed the previously reported positive effect of increased MOI on correction rates.

These data show that AAV vectors can correct different types of mutation at various chromosomal positions with high fidelity.

Sequence homology and mutation localization

Efficient recombination with conventional targeting vectors such as adenovirus or plasmid DNA occurs only if the viral vectors share at least 3–4 kb of sequence homology with the chromosomal target. The rAAV targeting system limits the total length of homology shared between the chromosomal target and the vector to 5 kb. However, it has been shown that, in the rAAV system, only 150 bases on the 5′ end (short arm of homology) and 500 bases on the 3′ end (long arm) are needed for efficient targeting[44]. However, an increase in total homology between the substrate and the rAAV vector from 1.7–3 kb led to a 5-fold increase in targeting frequency.

Moreover, positioning of the mutation to be corrected with respect to the rAAV genome has a significant effect on targeting rates. The highest targeting frequency was detected when the mutation was closest to the centre of the viral DNA. Correction rates of mutations located only 0–37 bp from the ITR were close to background levels[44]. Central positioning of the mutation might increase the chances of each ITR, known to be essential for gene targeting (see above), assisting in the targeting process.

Gene-targeting efficiency and DSBs

Conventional gene targeting with transfected, linear dsDNA constructs is thought to occur by DSB-induced HR. This was shown previously by the introduction of a DSB in the target sequence, which stimulated the frequencies of gene targeting by 100- to several 1,000-fold[12,58–60]. Furthermore, the inactivation of certain HR proteins reduced the rate of DSB-induced HR and conventional gene targeting[61,62].

Recently, two groups showed that the introduction of DSBs into the chromosomal target could further improve rAAV targeting efficiencies. They compared the correction rates of either a mutant *GFP* or *lacZ* gene in the presence or absence of a DSB that was induced in the target gene prior to, or simultaneous with, targeting[63,64]. Such DSBs resulted in up to 100-fold increased targeting frequencies. In one study, approximately one in five DSBs was repaired by rAAV, which might indicate a limit for gene targeting corresponding to the number of cells with DSBs at their target site[63]. In both studies, the frequency of random integration was not affected. Therefore, a DSB might shift the ratio of targeted events to random-integration events towards targeting. Taken together, these data indicate that the mechanism of rAAV-mediated gene targeting involves the repair of DSBs by HR independent of random integration by NHEJ.

Mechanism of rAAV-mediated gene targeting

Understanding the molecular basis for the impressive targeting frequencies obtained with rAAV vectors is important, as it could provide a first step towards manipulating the process to achieve high targeting rates with a lower rate of random integration. Studies of rAAV integration into induced DSBs[46] and its preferential integration into actively transcribed regions[26] led to the hypothesis that random rAAV integration is facilitated by NHEJ or a variation of NHEJ that uses microhomologies between the vector and the cellular
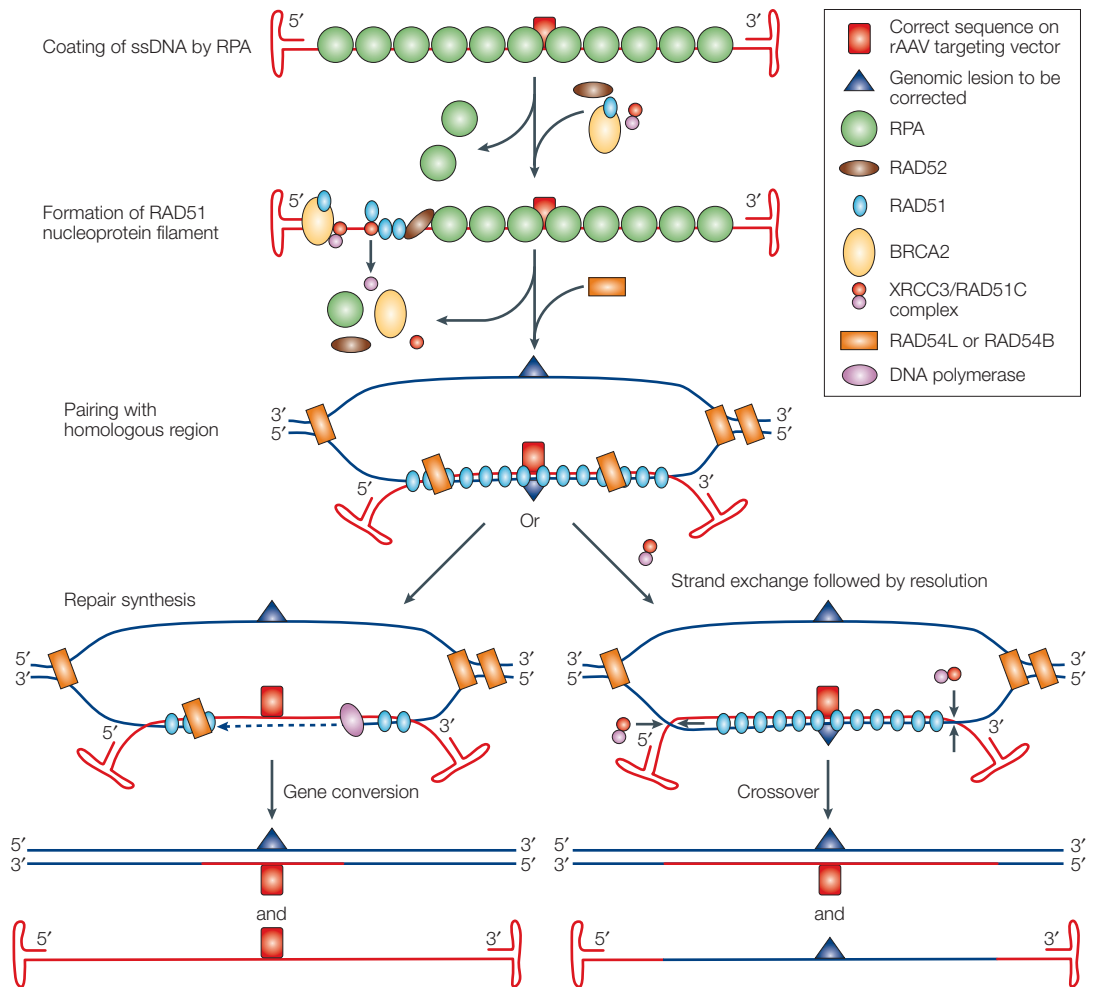
Figure 3 | **Molecular mechanism of rAAV-mediated gene targeting.** After removal of the viral capsid, replication protein A (RPA) binds to the single-stranded (ss) adeno-associated virus (AAV) genome and initiates homologous recombination. The BRCA2–RAD51 complex replaces RPA with the recombinase RAD51 with the assistance of RAD51 paralogues such as XRCC3/RAD51C. The resultant nucleoprotein filament finds and pairs with the homologous chromosomal sequence by unwinding and supercoiling that is facilitated by a protein known as RAD54L. Resolution of vector–chromosomal DNA intermediates results in introduction of the targeted modification in the homologous chromosome either by crossing over and exchange of sequences (lower panel, right) or by non-crossover gene conversion (lower panel, left). The recombinant AAV (rAAV) genome is shown in red and the chromosomal DNA in blue.

genome. In the presence of regions of homology between vector and target, another pathway takes place.

Recent experiments in the authors' laboratory, which used cell lines in which specific components of the HR or NHEJ pathways were silenced, indicate that the RAD51/RAD54 pathway of HR is required for rAAV targeting, whereas NHEJ does not contribute (Vasileva *et al.,* unpublished results). Based on the known features of HR proteins combined with the available data on rAAV-mediated targeting, we propose a molecular model for the mechanism of rAAV-mediated targeting (FIG. 3). Two studies have confirmed the importance of the ssDNA genome of parvoviruses for gene targeting[43,44] and suggested that the ITRs of rAAV, which resemble processed DNA breaks with a transition from ssDNA to dsDNA, might induce DNA-damage responses. Monomeric rAAV ssDNA is prone to degradation by nucleases[65],

which might be prevented by coating of the single-stranded region by a protein with high-affinity for ssDNA such as replication protein A (RPA). This heterotrimeric protein also prevents the formation of secondary DNA structures, but it is unknown, yet important to study, whether it affects the rAAV ITRs. HR is initiated by recruiting the BRCA2–RAD51 complex. Consistent with this, a BRCA2 homologue was shown to bind preferentially to a ssDNA–dsDNA transition that is identical to the end of the 5′-rAAV ITR, and it replaces RPA with RAD51 (REF. 66). RAD51 paralogues such as XRCC3 might be involved, as they play a part in HR[67,68], perhaps to support the formation of the RAD51 nucleoprotein filament on the ssDNA. The RAD51 filament then finds and pairs with homologous sequences in the genome, that is, the chromosomal locus to be repaired. This might require helicases like RAD54L and RAD54B,

which interact with RAD51 and unwind the target DNA strands. One way to resolve the resulting recombination intermediate is by GENE CONVERSION. This repair pathway is predominant in mammalian cells[69,70] and might be responsible for the increase in rAAV targeting following a DSB in the chromosomal target. DNA synthesis and extension of the broken chromosomal strand using the virus sequence as a template copies the correct sequence into the genome. Alternatively, Holliday junctions can be cleaved by resolvases, resulting in strand exchange between the virus vector and the cellular genome. In both cases, one strand of the target DNA can be repaired and, therefore, the correct sequence is transmitted to the next cell generation. Obviously, many aspects of this model need to be tested, such as the formation of a RAD51 filament, formation of paired structures and the resolution of intermediates.

Applications of rAAV gene targeting

The broad host range, the availability of multiple subtypes of the virus and recent advances in viral re-targeting to specific cell types by manipulation of capsid determinants[71,72] make rAAV an attractive system for gene correction and for gene disruption to generate knockouts in cultured human primary cells or cell lines (REFS 73–76 and P. Jallepali, personal communication).

Advances in the production of rAAV vectors that are devoid of contaminating replication-competent virus, and the high demand for effective therapies, have brought rAAV vectors into the clinical-research arena. Successful alleviation of severe hereditary disorders such as cystic fibrosis, haemophilia B, Duchenne muscular dystrophy[77], Parkinson's disease[78] and type I diabetes[79] in rodent and non-human primate models prompted Phase I and II trials in patients with haemophilia B or cystic fibrosis[80,81].

Although impressive, the gene-targeting frequency of up to 1% in cell lines is still accompanied by about 10% random integration, with its inherent risk. As any random rAAV integration event can lead to lethal or carcinogenic effects, it would be ideal to exclude such random events. Adult stem cells offer this possibility, as they have the potential to treat many diseases and can be subjected to *ex vivo* genetic manipulation prior to autologous transplantation. These cells can be transduced with the appropriate rAAV targeting vector, which modifies the mutated gene causing an illness. Subsequent selection enriches for cells that have undergone gene targeting, and Southern analysis of genomic DNA from expanded resistant clones selects out those in which random integration has occurred in parallel. Following expansion and analysis, the modified cells can be re-introduced into the patient.

Preliminary experiments using rAAV vectors in normal human fibroblasts yielded disruption of a gene that encodes a component of collagen fibres (*COL1A1*) at a maximum frequency of 1%. The efficient design of a 1.5-kb insert of an internal ribosome entry site–*neo* cassette allowed for selection and 70% enrichment for targeted clones that were devoid of random integration[57]. Dominant mutations in the *COL1A1* gene cause the brittle bone disease OSTEOGENESIS IMPERFECTA (OI) and, therefore, a disruption that generates a null allele should alleviate this disorder. The successful disruption of *COL1A1* in cell culture was followed by therapeutic targeting experiments in non-haematopoietic, pluripotent mesenchymal stem cells (MSCs)[82] (see Supplementary information S4 (figure)). These MSCs can be differentiated into the different cell types that are present in bone, cartilage, tendon, ligament, muscle and adipose tissue (reviewed in REF. 83). The transduction of MSCs isolated from young patients with moderately severe OI (type III) with an rAAV vector followed by neomycin selection yielded 90% targeting, and only a few clones displayed random integration. As confirmed by implants in mice, the polyclonal population and most of the targeted clones were capable of bone formation and showed improved stability and structure of the collagen fibres. This demonstration of successful rAAV-mediated gene targeting in adult human stem cells illustrates their potential in developing effective therapies.

Bovine spongiform encephalopathy (BSE) is a prion disease associated with the accumulation of an abnormal isoform of host-encoded prion protein (PrP) in the brain (reviewed in REF. 84). Elimination of this protein in mice can stop disease progression, and constitutive lack of PrP is well tolerated. Neutralizing this fatal neurodegenerative disease in cattle was attempted by Hirata and colleagues[85], who disrupted the normal *PRNP* gene (which encodes PrP) by rAAV targeting of embryonic bovine fibroblasts. Notably, none of the targeted clones contained random integrants, which makes the targeted cells perfect nuclear donors to generate BSE-resistant cloned cattle.

Conclusions and future perspectives

The studies reviewed here provide first insights into the potential of AAV-based targeted gene therapy; however, many aspects still remain unexplored. The rapidly expanding knowledge of the molecular basis of disease and technological improvements in vector design, cell differentiation and manipulation hold great promise for future applications of the unique properties of AAV vectors.

As the molecular basis for the high targeting frequencies obtained with rAAV vectors remains elusive, an understanding of its mechanism as hypothesized in FIG. 3 is important and might provide the next step towards manipulating the process and achieving high targeting frequencies with less simultaneous random integration. Learning the 'tricks' used by AAV to efficiently carry out HR might also provide groundwork for manipulation or adaptation of other gene targeting systems. The existing data not only point to promising future therapeutic applications, but also highlight a unique system that is of great interest to research that aims at understanding parvovirus biology and fundamental aspects of cell biology such as DNA recombination and repair.

GENE CONVERSION
A non-reciprocal transfer of genetic information.

OSTEOGENESIS IMPERFECTA
The most common and the mildest form of a group of rare disorders affecting the connective tissue, which are characterized by fragile bones that fracture easily (brittle bones).

1. Baum, C. *et al.* Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood* **101**, 2099–2114 (2003).

2. Lu, Y. Recombinant adeno-associated virus as delivery vector for gene therapy — a review. *Stem Cells Dev.* **13**, 133–145 (2004).

3. McCarty, D. M., Young, S. M. Jr. & Samulski, R. J. Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annu. Rev. Genet.* **38**, 819–845 (2004).

4. Hacein-Bey-Abina, S. *et al.* A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **348**, 255–256 (2003).

5. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).

6. Deng, C. & Capecchi, M. R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* **12**, 3365–3371 (1992).

7. Jasin, M. Chromosome breaks and genomic instability. *Cancer Invest.* **18**, 78–86 (2000).

8. Torres, R. M. & Kühn, R. *Laboratory Protocols for Conditional Gene Targeting* (Oxford Uninversity Press, Oxford, 1997).

9. McClarrinon, M. *et al. In vivo* studies of gene expression via transient transgenesis using lipid-DNA delivery. *DNA Cell Biol.* **18**, 533–547 (1999).

10. Xiang, J. J. *et al.* IONP-PLL: a novel non-viral vector for efficient gene delivery. *J. Gene Med.* **5**, 803–817 (2003).

11. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA* **91**, 6064–6068 (1994).

12. Brenneman, M., Gimble, F. S. & Wilson, J. H. Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases. *Proc. Natl Acad. Sci. USA* **93**, 3608–3612 (1996).

13. Bibikova, M. *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* **21**, 289–297 (2001).

14. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).

15. Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764 (2003).

16. Epinat, J. C. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* **31**, 2952–2962 (2003).

17. Fujita, A., Sakagami, K., Kanegae, Y., Saito, I. & Kobayashi, I. Gene targeting with a replication-defective adenovirus vector. *J. Virol.* **69**, 6180–6190 (1995).

18. Thevenot, E. *et al.* Targeting conditional gene modification into the serotonin neurons of the dorsal raphe nucleus by viral delivery of the Cre recombinase. *Mol. Cell. Neurosci.* **24**, 139–147 (2003).

19. Russell, D. W. & Hirata, R. K. Human gene targeting by viral vectors. *Nature Genet.* **18**, 325–330 (1998).
   The seminal paper that first showed high frequencies of gene targeting by rAAV.

20. Flierl, A., Chen, Y., Coskun, P. E., Samulski, R. J. & Wallace, D. C. Adeno-associated virus-mediated gene transfer of the heart/muscle adenine nucleotide translocator (ANT) in mouse. *Gene Ther.* **12**, 570–578 (2005).

21. Bertran, J. *et al.* Recombinant adeno-associated virus-mediated high-efficiency, transient expression of the murine cationic amino acid transporter (ecotropic retroviral receptor) permits stable transduction of human HeLa cells by ecotropic retroviral vectors. *J. Virol.* **70**, 6759–6766 (1996).

22. Nakai, H., Storm, T. A. & Kay, M. A. Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver *in vivo*. *J. Virol.* **74**, 9451–9463 (2000).

23. Nakai, H. *et al.* Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction *in vivo*. *J. Virol.* **75**, 6969–6976 (2001).

24. Nakai, H. *et al.* AAV serotype 2 vectors preferentially integrate in active genes in mice. *Nature Genet.* **34**, 297–302 (2003).

25. Nakai, H. *et al.* Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J. Virol.* **79**, 3606–3614 (2005).

26. Miller, D. G., Rutledge, E. A. & Russell, D. W. Chromosomal effects of adeno-associated virus vector integration. *Nature Genet.* **30**, 147–148 (2002).

27. Linden, R. M., Ward, P., Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus. *Proc. Natl Acad. Sci. USA* **93**, 11288–11294 (1996).

28. Samulski, R. J. *et al.* Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* **10**, 3941–3950 (1991).

29. Kotin, R. M., Linden, R. M. & Berns, K. I. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* **11**, 5071–5078 (1992).

30. Schmidt, M., Afione, S. & Kotin, R. M. Adeno-associated virus type 2 Rep78 induces apoptosis through caspase activation independently of p53. *J. Virol.* **74**, 9441–9450 (2000).

31. Tan, M., Qing, K., Zhou, S., Yoder, M. C. & Srivastava, A. Adeno-associated virus 2-mediated transduction and erythroid lineage-restricted long-term expression of the human β-globin gene in hematopoietic cells from homozygous β-thalassemic mice. *Mol. Ther.* **3**, 940–946 (2001).

32. Muzyczka, N. Use of adeno-associated virus as a general transduction vector for mammalian cells. *Curr. Top. Microbiol. Immunol.* **158**, 97–129 (1992).

33. Duan, D. *et al.* Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *J. Virol.* **72**, 8568–8577 (1998).

34. Yan, Z., Zhang, Y., Duan, D. & Engelhardt, J. F. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proc. Natl. Acad. Sci. USA* **97**, 6716–6721 (2000).

35. Duan, D., Yue, Y. & Engelhardt, J. F. Expanding AAV packaging capacity with trans-splicing or overlapping vectors: a quantitative comparison. *Mol. Ther.* **4**, 383–391 (2001).

36. Goncalves, M. A., Pau, M. G., de Vries, A. A. & Valerio, D. Generation of a high-capacity hybrid vector: packaging of recombinant adenoassociated virus replicative intermediates in adenovirus capsids overcomes the limited cloning capacity of adenoassociated virus vectors. *Virology* **288**, 236–246 (2001).

37. Chao, H., Sun, L., Bruce, A., Xiao, X. & Walsh, C. E. Expression of human factor VIII by splicing between dimerized AAV vectors. *Mol. Ther.* **5**, 716–722 (2002).

38. Reich, S. J. *et al.* Efficient trans-splicing in the retina expands the utility of adeno-associated virus as a vector for gene therapy. *Hum. Gene Ther.* **14**, 37–44 (2003).

39. Ellis, J. & Bernstein, A. Gene targeting with retroviral vectors: recombination by gene conversion into regions of nonhomology. *Mol. Cell. Biol.* **9**, 1621–1627 (1989).

40. Liu, X. *et al.* Targeted correction of single-base-pair mutations with adeno-associated virus vectors under nonselective conditions. *J. Virol.* **78**, 4165–4175 (2004).
   Compared gene targeting between the HEK293 cell line and primary mouse fibroblasts, revealing the importance of the natural AAV host. Narrowed down the occurrence of targeting to the S phase of the cell cycle.

41. Qing, K. *et al.* Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nature Med.* **5**, 71–77 (1999).

42. Hansen, J., Qing, K., Kwon, H. J., Mah, C. & Srivastava, A. Impaired intracellular trafficking of adeno-associated virus type 2 vectors limits efficient transduction of murine fibroblasts. *J. Virol.* **74**, 992–996 (2000).

43. Hendrie, P. C., Hirata, R. K. & Russell, D. W. Chromosomal integration and homologous gene targeting by replication-incompetent vectors based on the autonomous parvovirus minute virus of mice. *J. Virol.* **77**, 13136–13145 (2003).

44. Hirata, R. K. & Russell, D. W. Design and packaging of adeno-associated virus gene targeting vectors. *J. Virol.* **74**, 4612–4620 (2000).
   This study attempts to define the minimal stretch of homology necessary for efficient targeting and proves the importance of single-stranded rather than double-stranded vector genomes.

45. Russell, D. W., Miller, A. D. & Alexander, I. E. Adeno-associated virus vectors preferentially transduce cells in S phase. *Proc. Natl Acad. Sci. USA* **91**, 8915–8919 (1994).

46. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nature Genet.* **36**, 767–773 (2004).

47. Alexander, I. E., Russell, D. W. & Miller, A. D. DNA-damaging agents greatly increase the transduction of nondividing cells by adeno-associated virus vectors. *J. Virol.* **68**, 8282–8287 (1994).

48. Russell, D. W., Alexander, I. E. & Miller, A. D. DNA synthesis and topoisomerase inhibitors increase transduction by adeno-associated virus vectors. *Proc. Natl Acad. Sci. USA* **92**, 5719–5723 (1995).

49. Trobridge, G., Hirata, R. K. & Russell, D. W. Gene targeting by adeno-associated virus vectors is cell-cycle dependent. *Hum. Gene Ther.* **16**, 522–526 (2005).

50. Saleh-Gohari, N. & Helleday, T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res.* **32**, 3683–3688 (2004).

51. Yamamoto, A. *et al.* Cell cycle-dependent expression of the mouse *Rad51* gene in proliferating cells. *Mol. Gen. Genet.* **251**, 1–12 (1996).

52. Chen, F. *et al.* Cell cycle-dependent protein expression of mammalian homologs of yeast DNA double-strand break repair genes *Rad51* and *Rad52*. *Mutat. Res.* **384**, 205–211 (1997).

53. Hinz, J. M., Yamada, N. A., Salazar, E. P., Tebbs, R. S. & Thompson, L. H. Influence of double-strand-break repair pathways on radiosensitivity throughout the cell cycle in CHO cells. *DNA Repair (Amst.)* **4**, 782–792 (2005).

54. Inoue, N., Hirata, R. K. & Russell, D. W. High-fidelity correction of mutations at multiple chromosomal positions by adeno-associated virus vectors. *J. Virol.* **73**, 7376–7380 (1999).

55. Brown, P. O. Integration of retroviral DNA. *Curr. Top. Microbiol. Immunol.* **157**, 19–48 (1990).

56. Inoue, N., Dong, R., Hirata, R. K. & Russell, D. W. Introduction of single base substitutions at homologous chromosomal sequences by adeno-associated virus vectors. *Mol. Ther.* **3**, 526–530 (2001).
   Expands the types of mutation introduced by rAAV targeting vectors to single nucleotide substitutions, indicating the involvement of base-excision repair in the process.

57. Hirata, R., Chamberlain, J., Dong, R. & Russell, D. W. Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nature Biotechnol.* **20**, 735–738 (2002).
   Pilot study showing high frequency of *ex vivo* targeted disruption of a dominant negative allele of the *COL1A1* gene that causes OI.

58. Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.* **23**, 5012–5019 (1995).

59. Sargent, R. G., Brenneman, M. A. & Wilson, J. H. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell. Biol.* **17**, 267–277 (1997).

60. Taghian, D. G. & Nickoloff, J. A. Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells. *Mol. Cell. Biol.* **17**, 6386–6393 (1997).

61. Essers, J. *et al.* Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. *Cell* **89**, 195–204 (1997).

62. Rijkers, T. *et al.* Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. *Mol. Cell. Biol.* **18**, 6423–6429 (1998).

63. Miller, D. G., Petek, L. M. & Russell, D. W. Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol. Cell. Biol.* **23**, 3550–3557 (2003).
   This study shows the augmentation of gene-targeting frequencies by the introduction of a DSB in the genomic target.

64. Porteus, M. H., Cathomen, T., Weitzman, M. D. & Baltimore, D. Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol. Cell. Biol.* **23**, 3558–3565 (2003).
   Published back-to-back with reference 63, this study confirmed the enhancement of gene targeting by DSBs in live cells by monitoring *GFP* repair.

65. Hauck, B., Zhao, W., High, K. & Xiao, W. Intracellular viral processing, not single-stranded DNA accumulation, is crucial for recombinant adeno-associated virus transduction. *J. Virol.* **78**, 13678–13686 (2004).

66. Yang, H., Li, Q., Fan, J., Holloman, W. K. & Pavletich, N. P. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA–ssDNA junction. *Nature* **433**, 653–657 (2005).

67. Forget, A. L., Bennett, B. T. & Knight, K. L. Xrcc3 is recruited to DNA double strand breaks early and independent of Rad51. *J. Cell. Biochem.* **93**, 429–436 (2004).

68. Liu, Y., Masson, J. Y., Shah, R., O'Regan, P. & West, S. C. RAD51C is required for Holliday junction processing in mammalian cells. *Science* **303**, 243–246 (2004).

69. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19**, 3398–3407 (2000).

70. Johnson, R. D. & Jasin, M. Double-strand-break-induced homologous recombination in mammalian cells. *Biochem. Soc. Trans.* **29**, 196–201 (2001).

71. White, S. J. *et al.* Targeted gene delivery to vascular tissue *in vivo* by tropism-modified adeno-associated virus vectors. *Circulation* **109**, 513–519 (2004).

72. Ried, M. U., Girod, A., Leike, K., Buning, H. & Hallek, M. Adeno-associated virus capsids displaying immunoglobulin-binding domains permit antibody-mediated vector retargeting to specific cell surface receptors. *J. Virol.* **76**, 4559–4566 (2002).

73. Kohli, M., Rago, C., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. *Nucleic Acids Res.* **32**, e3 (2004).

74. Cummins, J. M. *et al.* Tumour suppression: disruption of *HAUSP* gene stabilizes p53. *Nature* **428**, 1 p following 486 (2004).

75. Cummins, J. M. *et al.* X-linked inhibitor of apoptosis protein (XIAP) is a nonredundant modulator of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-mediated apoptosis in human cancer cells. *Cancer Res.* **64**, 3006–3008 (2004).

76. Samuels, Y. *et al.* Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**, 561–573 (2005).

77. Goncalves, M. A. *et al.* Transfer of the full-length dystrophin-coding sequence into muscle cells by a dual high-capacity hybrid viral vector with site-specific integration ability. *J. Virol.* **79**, 3146–3162 (2005).

78. During, M. J., Kaplitt, M. G., Stern, M. B. & Eidelberg, D. Subthalamic GAD gene transfer in Parkinson disease patients who are candidates for deep brain stimulation. *Hum. Gene Ther.* **12**, 1589–1591 (2001).

79. Song, S. *et al.* Recombinant adeno-associated virus-mediated α-1 antitrypsin gene therapy prevents type I diabetes in NOD mice. *Gene Ther.* **11**, 181–186 (2004).

80. Manno, C. S. *et al.* AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. *Blood* **101**, 2963–2972 (2003).

81. Flotte, T. R. *et al.* Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult cystic fibrosis patients: a two-part clinical study. *Hum. Gene Ther.* **14**, 1079–1088 (2003).

82. Chamberlain, J. R. *et al.* Gene targeting in stem cells from individuals with osteogenesis imperfecta. *Science* **303**, 1198–1201 (2004).
    The closest to 'the clinic' that gene therapy with rAAV targeting vectors has reached to date.

83. Baksh, D., Song, L. & Tuan, R. S. Adult mesenchymal stem cells: characterization, differentiation, and application in cell and gene therapy. *J. Cell. Mol. Med.* **8**, 301–316 (2004).

84. Mallucci, G. & Collinge, J. Rational targeting for prion therapeutics. *Nature Rev. Neurosci.* **6**, 23–34 (2005).

85. Hirata, R. K. *et al.* Efficient *PRNP* gene targeting in bovine fibroblasts by adeno-associated virus vectors. *Cloning Stem Cells* **6**, 31–36 (2004).
    Shows the potential of the rAAV gene-targeting approach to generate cattle resistant to the prion disease BSE.

86. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).

87. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).

Online links

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez
AAV2 | COL1A1 | MUM | PRNP | Saccharomyces cerevisiae
Swiss-Prot: http://www.expasy.ch
Artemis | ATM kinase | BRCA2 | DNA ligase IV | ERCC1 | KU70 | KU86 | MBS85 | MRE11 | NBS1 | RAD50 | RAD51 | RAD51C | RAD52 | RAD54B | RAD54L | RPA | XPF | XRCC3 | XRCC4
CDC Infectious Disease Information: http://www.cdc.gov/ncidod/diseases/index.htm
BSE

FURTHER INFORMATION
Rolf Jessberger's homepage: http://www.genetherapy-mssm.net/bios/jessberger/jessberger.html

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (table) | S3 (table) | S4 (figure)
Access to this links box is available online.