Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/37412 holds various files of this Leiden University dissertation.

Author: Holkers, Maarten
Title: The roles of adenoviral vectors and donor DNA structures on genome editing
Issue Date: 2016-01-26

The roles of adenoviral vectors and donor DNA structures on genome editing

Maarten Holkers

The roles of adenoviral vectors and donor DNA structures on genome editing

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 26 januari 2016
klokke 15:00 uur

door

Maarten Holkers

geboren te Eibergen
in 1980

CONTENTS

Introduction

1

INTRODUCTION

Accurate and efficient genome editing is primarily dependent on the generation of a sequence-specific, genomic double-stranded DNA break (DSB) combined with the introduction of an exogenous DNA template into target cells. The exogenous template, called donor DNA, normally contains the foreign sequences flanked by DNA regions sharing sequence identity ("homologous") to those bracketing the target site. The strategies for mediating the formation of DSBs at the predefined genomic loci, have been undergoing intense investigation since the introduction in the mid 90's of sequence-customizable zinc-finger nuclease (ZFN) technology [1]. More recently, prokaryotic protein-based transcription activator-like effector nucleases (TALENs) [2] and RNA-guided nucleases (RGNs) derived from clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (Cas9) complexes [3, 4] have substantially broadened the availability and applicability of designer nuclease-mediated genome editing. Methods for the introduction of designer nucleases into target cells range from the plasmid transfection of cell lines with transfection reagents (e.g. polycations and liposomes) to the electroporation of haematopoietic stem cells with messenger RNA [5]. In general, however, the most effective methods for introducing foreign DNA into target cells are those based on viral vectors. The availability of diverse viral vector systems each of which with their own target cell tropisms, permit transducing a broad range of cell types. Importantly, viral vector particle-mediated gene delivery hands researchers with the possibility of controlling in a precise manner the copy-number of expression units reaching target cell nuclei and, as a result, the yields of encoded products. In this chapter, the main designer nuclease technologies are described along with the advantages and shortcomings associated with commonly used viral vector systems such as adeno-associated viral vectors (rAAV) [6], integrase-defective lentiviral vectors (IDLVs) [7] and adenoviral vectors (AdVs) [8].

Gene therapy

Correcting the phenotype associated with monogenetic disorders has been the focus of many researchers for many years. Historically, the first gene-therapy approaches were those based on the introduction of an exogenous correct copy of the mutated gene into target cells. Oncoretroviral vectors were generally used owing to their capacity of integrating its viral genome into host cell chromosomes. However, the strategy of random integration of an expression unit, flanked by unwarranted viral sequences, might prime serious adverse effects caused by insertional mutagenesis. Integration close to or within proto-oncogenes, might deregulate their expression in such a way as to induce monoclonal malignant cell expansion [9]. Next to this, the viral sequences flanking the expression unit are targets for silencing rendering, in most instances, these gene-therapy approaches ineffective [10].

Lastly, ectopic expression of gene products during early differentiation stages of stem cells can arrest these cells and eventually lead to cell death. Despite these major problems, a great deal of effort has been made into the design, optimisation and testing of viral vectors for the introduction of genetic payloads into target cells. In the first efficacious clinical trials, integration-proficient oncoretroviral vectors were used in order to integrate an expression unit in human long-term repopulating haematopoietic stem cells, relieving the X-linked severe combined immunodeficiency disorder (SCID-X1) in boys [11 - 13]. SCID-X1 is characterised by a block in T-cell differentiation combined with an impairment in B-cell immunity and is lethal during the first years of life. X-linked SCID is linked to mutations in the interleukin 2 receptor gamma gene *IL2-RG* whose product is responsible for cytokine-dependent signalling leading to the proliferation of T-cells. Adenosine deaminase (ADA)-SCID has also been shown to be amenable to oncoretroviral vector-based gene therapy protocols. This monogenetic disorder originates from the lack of functional ADA, which leads to a reduced lymphocyte proliferation. The successes of some of these gene-therapy clinical trials were, however, clouded by the onset of leukaemia in several of SCID-X1 patients. [9, 14]. Extensive characterisation of the vector integration profile in the patient cells revealed the underlying problems associated with these first-generation oncoretroviral vectors, namely, viral enhancer-driven proto-oncogene deregulation (e.g. *LMO2*) [15-18]. Following these clinical results, the design of these vectors has been improved by the removal of most of the viral promoter elements and cryptic splicing sequences combined with the use of less powerful cellular promoters to drive therapeutic transgene expression [19, 20]. In addition, new retroviral vectors based on lentiviruses (e.g. human immunodeficiency virus type 1), as opposed to murine oncoretroviral vectors, have also contributed to a further improvement in gene-therapy protocols. These developments, in combination with extensive monitoring of the treated patients, paved the way for several, so far mostly uneventful follow-up clinical trials. Taken together, these clinical trials have provided the medical field with a genetic approach for treating monogenetic disorders, which has been more recently expanded to other conditions like metachromatic leukodystrophy and Wiskott-Aldrich syndrome. [21, 22].

Despite these steady improvements in clinical gene-therapy protocols, the preferred path to tackle monogenetic disorders would be to directly repair the disease-causing mutations without leaving a genetic "footprint" in the genome. Importantly, such genetic interventions should result in physiologically regulated gene expression without altering the expression of neighbouring genes. In order to introduce foreign DNA sequences into target cell genomes, the cellular DNA repair mechanisms can be exploited. Naturally, the most often used DNA repair pathway in mammalian cells is that of non-homologous end-joining (NHEJ) which, in essence, ligates the two broken chromosomal DNA ends back together often inserting

at the newly formed junction so-called small insertions and deletions (indels) [23]. Homologous recombination (HR) is an alternative DNA repair pathway in which sequences from a homologous chromosome or sister chromatid serve as templates for restoring broken chromosomal DNA. In the presence of exogenous DNA templates, this pathway can be exploited for introducing in a precise manner foreign DNA into specific positions in the genome. However, the efficiency of "spontaneous" HR between a target chromosomal region and a corrective donor-DNA fragment is very low, often described as a 'one-in-a-million' type of event, with efficiencies ranging from 10^{-8} to 10^{-6} [24, 25].

Importantly, these very low rates can be elevated substantially by actively triggering the endogenous DNA repair machineries. Pioneering experiments involving a homing endonuclease (HE) derived from the yeast *Saccharomyces cerevisiae* (I-SceI), provided a rationale for using site-specific nucleases as HR-mediated gene targeting triggers [26 - 28]. Indeed, the I-SceI HE (also known as 'meganuclease') recognises a specific 18 bp-long sequence and, after binding, generates a DSB provoking the cell to repair the lesion and, in doing so, increase the rate of HR-mediated gene insertion by multiple orders of magnitude. The human genome does not contain the consensus I-SceI target site, nevertheless, there are multiple cryptic target sites leading to I-SceI mediated DNA cleavage [29]. As the location of these ambiguous binding sites are unclear, other methods were embraced to generate site-specific DSBs into the human genome, preferentially into genomic "safe-harbour" regions like the *AAVS1* locus in the first intron of the *PPP1R12C* gene [30]. This region can be modified or used to integrate exogenous expression units without affecting the expression of neighbouring genes [31, 32]. Protein engineering of I-SceI and of other meganucleases did result in variations capable of identifying additional sequences occasionally existing in the human genome [33, 34]. Nonetheless, as the accuracy of sequence recognition of these proteins are not very strict as many "cleavable" pseudo-sites were identified by using rAAV DNA "trapping" [29]. Lastly, diverting these HRs to target other sequences than their natural target involves complex protein engineering which is a technique only available to a few research groups. Other type of protein engineering has been pushing the field of genome editing forward in the form of artificial site-specific nucleases based on zinc-finger motifs.

Zinc-finger nucleases

ZFNs are chimeric proteins combining the sequence-specific protein-DNA interaction properties of arrayed zinc-finger motifs, with the cleaving activity of dimeric type IIS endonucleases [1, 35, 36]. Each zinc-finger motif can be designed to recognize a specific nucleotide triplet. Thus, a combination of 3 motifs in a tandem array can potentially generate a DNA-binding protein recognizing a 9 nucleotide-long target DNA sequence. The array of zinc-finger motifs is fused via a flexible peptide linker to the nuclease domain of,

most commonly, the type IIS restriction enzyme FokI. The catalytic activity of FokI nuclease domains is dependent on their dimerization in order to create an active nuclease complex capable of generating a DSB. To enable this dimerization process, a second ZFN is co-delivered to specifically bind a sequence on the opposite DNA strand located in the vicinity of that bound by its partner. Hence, these chimeric proteins assemble at their target sites to form a dimeric complex separated by 6- to 10-bp spacer sequences allowing for FokI nuclease domain dimerization and subsequent site-specific digestion of the DNA [1]. Although ZFN-mediated DNA cutting depends on binding to two properly spaced 9 bp half-target sites (conventionally described as "left-hand" and "right-hand"), and mismatches disturb the binding efficiency of the zinc finger motifs, off-target activity has been documented and studied by many research groups [37 - 40]. This off-target activity can be reduced by including a 4th zinc finger motif in each ZFN whereas extending it even further has the trade-off of diminishing the overall activity [41]. Mutants of the FokI nuclease domain have been identified that, together, generate so-called forced heterodimeric FokI nuclease domains. These have been shown to reduce off-target DNA digestion owing to the fact that binding of the same ZFN pair member at off-target sites (i.e. homodimer assembly) is, to a large extent, prevented.

ZFNs have been adopted by numerous investigators as valuable tools for the generation of site-specific DSBs for research purposes. This research resulted in a wide variety of insights into DNA repair mechanisms, donor DNA requirements and also led to the generation of various cell lines and animal models [42-45]. Currently, the success of ZFN technology is best illustrated by the initiation of the first human clinical trial involving the use of designer nucleases. This trial (coded: SB-728) consists of using adenoviral vector (AdV)-delivered ZFNs to knockout *CCR5* alleles and, as consequence, generating HIV-resistant human CD4+ T-cells, to halt or perhaps reverse AIDS progression. On the other hand, the major drawbacks of the ZFN technology are their relatively low genomic target sequence coverage as well as the difficulty in designing and generating efficient ZFN pairs, which simultaneously, display low cytotoxicity. The relatively complex protein engineering platforms and extensive validation required for workable ZFN pair isolation continues to make it difficult the use of this technology by non-experts. In addition, the ZFN technology is comprehensively patented with only one ZFN scaffolding platform being commercially available for research purposes [46, 47].

Transcription activator-like effector nucleases
The aforementioned limitations related to ZFN-based genome editing approaches were partially lifted after the introduction of a new class of chimeric nucleases whereby the DNA recognition is attained by transcription activator-like effector (TALE) protein elements [48-50]. TALEs consist of a family of repeat-containing proteins derived from *Xanthomonas*

bacteria that, upon infection of host plant cells, act as virulence-enhancing transcriptional factor activators. Naturally, TALEs are organized into three distinct domains, the N-terminal region used for protein translocation to the nucleus, in order for binding to their DNA targets, the central DNA-binding domain containing a repeat array with repeat variable di-residues (RVDs) and a C-terminal region containing a nuclear localization signal sequence and the transcription activation domain. TALE proteins use a one-to-one protein-to-DNA code whereby the amino acids at positions 12 and 13 (i.e. RVDs) within each repeat of typically 34 amino acids define binding to a specific nucleotide. Thus, RVD-containing TALE repeats can be placed in tandem creating a sequence-specific DNA binding motif, which in turn can be coupled to the FokI nuclease domain to generate a TALE-nuclease (TALEN) monomer [51, 52]. In common with ZFNs, these designer proteins work in tandem whereby a pair of TALENs assemble at their respective half-target sites on opposite DNA strands separated by an intervening spacer sequence. In the case of TALENs, however, the spacer sequences are larger than those of ZFNs, typically between 14- and 20-bp long. Importantly, the reported cytotoxicity profile of TALENs, presumably caused by off-target effects, is significantly lower than that of ZFNs, making TALENs a potentially safer artificial nuclease system [53]. Although it was initially reported that the affinity of the TALE repeat is restricted to the binding of the N-terminal domain to a thymidine, more recent findings indicate that this is far from being a strict requirement making the theoretical target-site coverage of TALENs to every two to three nucleotides. The robustness and simplicity of the TALEN technology is permitting the broad use of these nucleases for genome editing purposes. This is further aided by the development of 'do-it-yourself' TALEN assembly systems [54-57] and by the availability of various commercial services for the outsourcing of TALEN pair generation.

RNA-guided nucleases

The two previously described nuclease systems depend on protein-DNA interactions for their target site selectivity. In contrast, RNA-guided nuclease (RGN) systems rely on the guiding of a non-specific nuclease by means of a single-stranded RNA corresponding to the target DNA sequence. RGNs are based on type II clustered regularly interspaced short palindromic repeats (CRISPR) innate immune systems found in bacteria and archaea. These innate immunity-related genetic systems evolved to counteract invading foreign DNA like that of plasmids and bacteriophages and are typically composed of a nuclease (Cas9) and short RNA moieties (CRISPR and trans-activating RNAs) The complex formed by Cas9, CRISPR RNA and trans-activating RNA scans the DNA for the presence of a so-called protospacer adjacent motif (PAM) of 3 bp (NGG) next to a CRISPR RNA-complementary sequence of approximately 20 bp. After binding to this target site, the tripartite complex cleaves the DNA strands leaving a blunt-ended DSB. The adaptation of this system to mammalian cells in the form of a RGN platform was independently published by several groups in early 2013

[3, 4, 58]. Of note, by fusing the tracrRNA to the CRISPR RNA, a functional single-guide RNA (gRNA) module can be formed hence simplifying the system and making it possible to change the specificity of the RGN complex by merely changing the 20 bases responsible for the recognition of the target site. As a result RGNs are easy to be redirected to new single, or to multiple, chromosomal target sites. Transfection of gRNA oligoribonucleotides or gRNA-encoding RNA polymerase III-based expression units can be combined with Cas9 delivery methods to the induced site-specific DSBs in target cells. Nevertheless, several publications reported the significant off-target cleaving activity of RGNs at DNA sequences containing up to 6 mismatches [59, 60]. Fu and co-workers searched *in silico* for the top 20 off-target sites for 3 gRNAs and analysed these sites for the presence of CRISPR/Cas9-dependent indels generated by NHEJ-mediated DNA repair. Independently of cell type or gRNA sequence, almost every predicted site contained T7 endonucleaseI (T7EI)-sensitive PCR products, indicative for indel formation at the predicted off-target sites [59 - 61]. Since these initial studies characterizing RGN off-target activity, considerable efforts have been undertaken to improve RGN specificity. These efforts included testing Cas9 proteins derived from different organisms, shortening the gRNA by 1 to 3 nucleotides (truncated-gRNA) to reportedly optimize thermodynamically related RNA-DNA hybridization parameters or by using pairs of Cas9 "nickase" mutants to induce single-stranded DNA breaks (nicks) in opposing DNA strands at the target site and, by doing so, generate a coordinated DSB [62, 63]. However, although the use of the latter 'double-nicking' approach does reduce the generation of NHEJ-mediated chromosomal DNA disruptions at off-target sites, there is still the risk that off-target nicks can lead to DSBs, such as, by an advancing replisome [64].

More recently, it has been published that catalytically inactive forms of Cas9 can be fused to a FokI domain and, after dimerization, generate site-specific DSBs at the target site. This approach of using the multiplexing properties of RGNs combined with the dimerization requirement of FokI domains, reduced the off-target effect of RGNs dramatically albeit, at the expense of generating rather bulky proteins [65, 66].

Delivery of designer nucleases into target cells

Non-viral delivery systems

Methods to introduce exogenous expression units into target cells are a key requirement for many research areas. These methodologies include transfecting plasmid DNA molecules into cells in culture by conjugating them to readily available cationic polymers or by encapsulating them in liposome vesicles [51, 52, 67]. These nucleic acid transfection protocols, however, are often not effective in primary cell cultures in part due to the slower division rates of these cells when compared to those of transformed cell lines. As the nuclease expression level is a key determinant for effective genome editing, improved methods to introduce these, often large, exogenous proteins into target cells are in demand. One way of delivering foreign genetic material into the cell nucleus is electroporation, whereby a short burst of current transiently opens the cell and nuclear membranes enabling the plasmid DNA or mRNA entry [68, 69]. This method, however, requires careful optimization on a per cell type basis and causes significant rates of cell-death. Considerable efforts have been made aiming at reducing the cell mortality and increasing the efficiency. These efforts resulted in protocols capable of transfecting human haematopoietic stem cells with ZFN mRNA displaying a measurable rate of gene disruption and the site-specific addition of an expression unit at the ZNF target sequences. Finally, zinc-finger motifs in ZFNs have a net positive charge and are therefore able to penetrate the cell surface of target cells [70]. TALENs lack this property, but were shown to be able to transduce target cells by fusing them to a poly-arginine protein transduction domain, also called cell-penetrating peptide (CPP). TALE repeats contain 18 solvent-exposed cysteine residues that can be conjugated to a thiol-reactive nitropyridyl (Npys) Arg9 peptide(R9) enabling cellular membrane penetration. As this linkage is reversible under reducing conditions, the CPP moiety is released from the TALENs by disulphide bond reduction following their cytosolic entry. Albeit at low levels, the resulting TALENs were shown to induce site-specific DNA cleavage [71].

Viral vectors

In general, the most efficient methods for introducing genetic material into a broad variety of target cells are those based on viral vectors. Viral vector development harnesses the natural ability of viruses to deliver their genetic material into host cells. Indeed, viral vectors have been enabling scientists to introduce foreign genetic payloads into a wide range of target cells, including primary human cells *in vitro* and *in vivo*. By replacing essential gene sequences of the parental virus by exogenous expression units, viral vectors can be made replication-defective. In general, viral vectors can be divided in integrating and non-integrating systems whereby the former class ensures stable transduction of dividing cells maintaining transgene expression, whereas the latter allows for transient expression

of specific proteins. Integration-competent murine oncoretroviral vector systems, such as that based on the murine Moloney leukaemia virus, have dominated the gene therapy research field for many years for the delivery of ectopic expression of transgenes. However, as aforementioned, the chromosomal DNA integration pattern of these oncoretroviral vectors is biased towards transcription start sites and CpG islands [72], increasing the risk of insertional mutagenesis by "landing" in or near a proto-oncogene like *LMO2* [9]. Lentiviral vector systems, on the other hand, are biased to integrate their proviral DNA into the coding sequences of transcriptionally active genes in both dividing and non-dividing cells [73, 74]. These characteristics, combined with the removal of promoter/enhancer elements from the long terminal repeats, so-called self-inactivating (SIN) versions, leads to a much-reduced genotoxicity profile [74]. Lentiviral vectors can be used to deliver designer nucleases into target cells with high efficiency. Nevertheless, as the prolonged expression of a sequence-targeting nuclease is not desired, the use of integration defective lentiviral vectors (IDLV) particles has been preferred. The main shortcoming of this non-integrating viral vector system is its very low expression level as a result of epigenetic silencing processes targeting episomal vector DNA [75, 76]. This silencing could, however, be ameliorated by using various histone deacetylase inhibitors [76, 77]. Importantly, lentiviral vectors deliver single-stranded positive-sense RNA molecules into target cells, which need first to be reverse transcribed into DNA by the vector particle-bound reverse transcriptase (RT). The low template affinity of RT combined with its RNaseH activity contributes to template switching events. This process is exacerbated when vector genomes contain regions sharing a high degree of sequence identity [78, 79]. For instance, the co-delivery of two ZFN ORFs in a single IDLV particle driven from a single promoter to yield an active nuclease complex from a single IDLV genome [77]. TALEN ORFs also suffer from this RT template switching phenomenon as shown by the precise deletion of several (up to 12) repeats from the TALE repeat region encoding the protein's DNA-binding domain [80, 81, This thesis]. Recently, changing the DNA sequence corresponding to the various TALE repeats by removing as much as possible repetitive sequences, established the ability of lentiviral vectors to introduce intact TALEN proteins in target cells. [82]. In a study by Cai and colleagues, the use of lentiviral vector particles as protein transducing carriers of designer nucleases was described [83]. In this work, the ZFN and TALEN proteins were fused to the lentiviral Gag precursors and, during virion maturation, the nucleases were released from Gag in the virus particles in a protease-dependent manner and found to be able to facilitate efficient DNA cleavage in transduced cells [83].

Recombinant AAV (rAAV) vectors contain a 4.7 kb single-stranded DNA genome which is flanked by two 145 bp inverted terminal repeats (ITRs). These palindromic, hairpin-forming, ITR sequences contain the viral origins of replication and packaging signals. This makes

them suitable for the delivery of ZFN ORFs and spacious enough to accommodate the 4.1-kb Cas9 nuclease gene from *S. pyogenes* [84, 85]. The lifecycle of the wild-type virus is divided in lytic and latent phases as a result of the presence and absence of a helper virus (e.g. AdV), respectively. In the latter case, the virus integrates in cultured cells its genome preferentially in a defined region of the human chromosome 19 called AAV integration site 1 (*AAVS1*). The integration reaction depends on the generation of a nick at a specific *AAVS1* target site that is mediated by the viral Rep78/68 proteins. In the absence of these proteins, such as following viral gene-deleted rAAV transductions, the genomes remain, for the most part, in an episomal state. In addition AAV transductions often result in a low expression of the transgene due, amongst other factors, to their single-stranded DNA nature [86, 87]. To counteract this low transgene expression, self-complementary rAAV (scAAV) systems have been developed. Following uncoating, the self-complementarity of scAAV genomes allows for the direct assembly of transcriptionally competent double-stranded DNA templates through intramolecular hybridization [88]. Indeed, the expression of transgenes delivered via scAAVs is higher when compared to that achieved by their conventional rAAVs counterparts. However, their halved cloning capacity (i.e. ~ 2.4 kb) makes them unsuitable for TALEN or Cas9 delivery.

AdVs have linear double-stranded DNA genomes packaged in a non-enveloped icosahedral capsid. To date, 57 human serotypes have been isolated. These can be divided in 7 species (A through G) and are found to cause a wide range of illnesses, ranging from mild respiratory diseases in children and the elderly, to life-threatening infections in people with a weakened immune system [89, 90]. AdVs are a commonly used vector platform whereby the deletion of several essential gene functions, usually, early (E) regions 1A (*E1A*) and 1B (*E1B*), often combined with the deletion of *E3*, creates space for the insertion of relatively large transgenes such as those encoding TALEN monomers and the *S. pyogenes* Cas9/CRISPR nuclease. Importantly, the deletion of the *E1* ORFs, renders these vectors replication incompetent in target cells and thus requires the use of *E1A*- and *E1B*-complementing cell lines like HEK293 [91], 911 [92] or PER.C6 [93] to provide *in trans* the missing functions during their production [8]. These *E1*-deleted or first-generation AdVs can be easily generated by using pAdEasy-based platforms [94, 95]. Although second-generation AdVs, combining the deletions of *E1A* and *E1B* with that of *E2A* or *E4*, can accommodate larger expression units, there are more constrains in their production since it requires the use of more specialised cell lines expressing the cytotoxic *E2A* or *E4* gene products [96]. The latest generation of AdVs, the so-called "gutless" or high-capacity AdVs, can accommodate up to 37 kb of foreign DNA since, from the parental virus genome, these vectors only retain the *cis*-acting AdV TRs and packaging signal. The production of these vectors is executed in the presence of a helper-vector to provide *in trans* all of the required AdV gene products. In order to remove

the co-replicating helper vector genomes from the packaged viral vector DNA fraction, the packaging signal of the helper is flanked by recombinase recognition sequences, for example loxP or FRT sites, and vector production takes place in complementing cell lines expressing the respective recombinase protein, Cre or FLP [97]. After several rounds of vector propagation, titers as high as 1×10^{12} vector particles per millilitre can be achieved.

The availability of different AdV serotypes permits constructing AdVs with the capacity to transduce a wide variety of target cells independently of their cell division status. The AdV genome doesn't integrate in the target cell genome, providing researchers with a vector system for the transient expression of designer nucleases. Owing to these properties, all the previously described classes of designer nucleases can be introduced into target cells by AdV systems in an efficient and controlled manner [80, 81, 99, This thesis].

Donor DNA templates

The activation of the cellular NHEJ DNA repair machinery in order to mutate genes via the introduction of a sequence-specific DSB is one of the approaches for editing the genome of a target cell. As aforementioned, the induction of a specific DNA repair mechanism depends, amongst other factors, on the cell-cycle status and on the presence of a donor DNA fragment to serve as a template to be inserted. The free-ended DNA end can, via strand-invasion, displace the target homologous sequence and, via base-pairing, use the complementary strand as template to copy the missing nucleotides. After completion, the generated DNA strand can flip back and be re-ligated to the broken strand completing this error-free DNA repair mechanism. Exogenously added DNA templates can be co-opted by the endogenous HR machinery to introduce in a targeted manner new genetic information into the target cell's chromosomal DNA. Such exogenous DNA molecules (donor DNA) can harbour, for instance, an expression unit or sequences for the correction of genetic mutations flanked by sequences homologous to those framing a designer nuclease's target site. Ideally, donor DNA templates are deprived of prokaryotic sequences and are on themselves non-integrating entities, nonetheless, can be used by the HR DNA repair mechanism to serve as templates for site-specific chromosomal DNA insertion. To date, the introduction of various donor DNA molecules in target cells has been predominantly mediated by transfection protocols for single-stranded oligonucleotides and plasmid DNA or, more prevailing in hard-to-transfect primary cell cultures, viral vectors like rAAVs and IDLVs. Despite the fact that short single-stranded DNA oligonucleotides have been shown to direct genetic correction *per se*, the combination of these molecules with the generation of a sequence-specific DSBs greatly enhances their use to fix the DNA lesion [100, 101]. Unquestionably, the use of these molecules greatly reduces the payload to be inserted impeding, for instance, the insertion of expression units. However, when used to correct single point mutations in open reading frames, they are most suited. Plasmid DNA molecules, when transferred into target cells,

can be adopted by the HR DNA repair machinery [102]. However, this methodology can, occasionally, co-integrate undesired flanking prokaryotic DNA sequences [103, This thesis]. Probably owing to the circular nature of plasmids, after copying the corrective nucleotides and likely due to the lack of downstream homologous sequences, the repair machinery presumably fails to switch back from the plasmid template to the endogenous genomic sequences, prompting the co-insertion of the plasmid backbone before completing the HR event. *In vitro* linearization of the plasmid DNA with restriction endonucleases can overcome this phenomenon, with the trade-off of reduced transfection efficiency in combination with the stimulation of exonuclease-mediated retention of the donor DNA sequences. Minicircle plasmids, deleted of prokaryotic sequences, prevent the integration of undesired methylation-prone prokaryotic elements, nevertheless their circular topology should not abolish the aforementioned "rolling-circle replication"-mediated type of insertions leading to undesired multiple copies [103, This thesis]. In principle, these illegitimate insertions pose no problems in gene-complementation protocols, albeit the insertions of multiple copies of the transgene can, in principle, yield deregulated transcriptional levels.

IDLV-mediated delivery of donor sequences benefits from the wide tropism of this vector-type, capable of transducing a wide variety of cells and has been successfully used as donor-DNA templates in HR-mediated gene conversion experiments [104], even, in combination with electroporation of ZFN mRNA molecules, in clinically relevant human haematopoietic stem cells [5]. Nonetheless, IDLV vectors suffer from several limitations. First, lentiviral vector genomes have a limited cloning capacity, which constrains the introducing of large targeting expression units. Moreover, intronic regions within these expression units or cryptic splice sites should ideally be excluded, as to prevent IDLV-driven aberrant RNA splicing. Related to this, such integrated IDLV genomes could cause alternative splicing, generating aberrant transcripts [105]. Next to this, the high error rates resulting from reverse transcription can generate *de novo* mutations [106]. Finally, the DNA structure of a fraction of the IDLV genomes resembles plasmid DNA molecules in that they display a so-called 1-LTR or a 2-LTR circular topology [7]. The circular templates build-up, next to the free-ended linear cDNA forms, presumably via NHEJ-mediated intramolecular recombination. The circular structures can mediate concatemeric insertions of multiple expression units flanked by undesirable viral elements, whereas the linear form has the tendency to home to naturally occurring chromosomal DSBs and become trapped at off-target sites. Despite this, IDLVs are widely used to deliver DNA donor templates into target cells. This preference is largely explained by that fact that high levels of stably transduced cells can be achieved by using IDLV donor DNA after nuclease- induced HR. However, this comes with the trade-off that a significant subset of stably transduced cells contain random and/or concatemeric integrants.

Conventional rAAVs and scAAVs are also used as donor DNA templates for site-specific integration, even in the context of gene targeting protocols that do not involve the generation of a DNA lesion [107]. Like IDLVs, rAAVs can be produced without any viral genes and have a wide range of tropisms making them good candidates for donor DNA template delivery. Their cloning capacity is, however, limited to approximately 4.7 kb. Next to this, off-target integrations in spontaneously occurring chromosomal DNA breaks generally dominate the overall integration profile of this vector type, often yielding 10 off-target integration events for each on-target integration [34].

The consequence of the use of sub-optimal donor DNA templates (i.e. templates with a promiscuous integration profile) for gene editing protocols led to the introduction of fail-safe elements designed to either remove the unsolicited integrations or to directly select the cells containing correctly integrated donor DNA by means of drug-based selection schemes. In the former case, the placement of, for example, a herpes simplex virus thymidine kinase (*HSV-TK*) expression unit "outside" the homologous sequences enables a selection method based on ganciclovir, which becomes a toxic product in cells containing the *HSV-TK* unit [108, 109], due to illegitimate chromosomal DNA insertion. In the latter case, a neomycin expression unit, for instance, can be placed between the flanking homologous sequenced in a intronic region, forcing the insertion of the donor template and ablating cells lacking neomycin [110]. These procedures depend, next to the introduction of sizable gene-expression cassettes, on very stringent selection protocols undesirable in most clinical genome modification protocols. An alternative method for the removal of unwanted donor template insertions, which does not depend on drug selections, is based on the excision of the exogenous integrated sequenced mediated by sequence-specific recombinases. The FLP and Cre recombinases can recognise their corresponding target-sites and if these are placed in a direct repeat orientation, mediate removal of the intervening DNA sequences [111-113]. This method will, however, leave a "footprint" in the genomic DNA in the form of one copy of the target site. Lastly, site-specific recombination events involving target-sites located on two different chromosomes might generate translocations, possibly prompting carcinogenesis [114, 115].

AdV vectors, as mentioned before, retain the ability to can transduce a wide variety of cell types and importantly, and accommodate relatively large genetic payloads. In the case of the previously described "gutless" AdV platform, this cloning capacity is as high as 37 kb. Still, up until recently, the use of this vector type as donor DNA delivery vehicles has been somewhat limited. In this regard, AdV-mediated donor DNA transfer has mostly involved gene targeting or gene repair strategies reliant on spontaneous HR and selection protocols. Nevertheless, after selection of the stably transduced cells, high rates of accurate integration at the desired

locations have been obtained [116, 117]. The target site-specificity and fidelity resulting from designer nuclease-induced AdV donor DNA insertion was only recently investigated [103, This thesis]. Experiments comparing this viral vector donor DNA type with that of IDLVs and plasmid DNA in the context of gene targeting models involving ZFNs, TALENs and RGNs revealed that, although the overall stable transduction efficiency of IDLV donor DNA was the highest among the different donor-DNA templates examined, the accuracy and specificity of the insertion of a transgene cassette at the desired locus was highest for the AdV donor DNA [103, This thesis]. The examined plasmid-based templates, similarly to IDLV donor DNA, showed a promiscuous integration pattern resulting in the integration of prokaryotic DNA sequences in addition to the off-target insertions and concatemeric forms of the exogenous DNA templates. By assessing different cell types and classes of designer nucleases, these gene targeting experiments established the unparalleled accuracy and specificity resulting from using AdV-delivered donor DNA templates into target cells. Validation of the hypothesis that the protein-capped DNA of AdV genomes formed the basis for this unsurpassed specificity and accuracy was achieved by a TALEN-induced *in vivo* digestion experiment, in which the release of the donor DNA template from the AdV genome readily led to illegitimate DNA integration. High-fidelity genetic modification of human cells, brought about by the delivery of both engineered nucleases and protein-capped donor DNA templates, has important implications to both fundamental and applied research protocols as it allows for accurate gene editing approaches without a strict requirement for using stringent selection regimes intended to remove donor DNA integrated at random genomic positions and/or as undefined concatemeric forms [103, This thesis].

To date, numerous efforts have been made to generate highly specific designer nucleases in order to mediate site-specific DNA lesions with a minimal off-target digestion profile. By fine-tuning ZFN target site selectivity, off-target effects can minimized, nevertheless, the cytotoxicity of these proteins still poses a problem. TALENs on the other hand are highly repetitive proteins exhibiting unsurpassed cleavage efficiency, combined with high sequence specificity and low cytotoxicity. The constant evolution in CRISPR protein and gRNA design, relieved the initially reported promiscuous properties of this popular genetic scissor. However, hitherto, the pursuit and optimization of high-fidelity donor DNA templates did not get the same emphasis as that devoted to improving the specificity of designer nuclease platforms. The finding that AdV-delivered protein-capped DNA donor templates results in highly specific and accurate genome modification following designer nuclease-induced DSBs might contribute to the on-going efforts in translating genome-editing concepts into novel human gene therapies.

REFERENCES

1. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci USA. 93, 1156-60 (1996)

2. Christian, M. *et al*. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 186, 757-761 (2010)

3. Cong, L. *et al*. Multiplex genome engineering using CRISPR/Cas systems. Science. 339, 819-823 (2013)

4. Mali, P. *et al*. RNA-guided human genome engineering via Cas9. Science. 339, 823-826 (2013)

5. Genovese, P. *et al*. Targeted genome editing in human repopulating haematopoietic stem cells. Nature 510, 235-240 (2014)

6. Russell, D.W. & Hirata, R.K. Human gene targeting by viral vectors. Nat Genet. 18, 325-330 (1998)

7. Wanisch, K. & Yáñez-Muñoz, R.J. Integration-deficient lentiviral vectors: a slow coming of age. Mol Ther. 17, 1316-1332 (2009)

8. Gonçalves, M.A. & de Vries, A.A. Adenovirus: from foe to friend. Rev Med Virol. 16, 167-186 (2006)

9. Hacein-Bey-Abina, S. *et al*. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science. 302, 415-419 (2003)

10. Ellis, J. Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum Gene Ther. 16, 1241-1246 (2005)

11. Cavazzana-Calvo, M. *et al*. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science. 288, 669-672 (2000)

12. Hacein-Bey, S., Yates, F., de Villartay, J.P., Fischer, A. & Cavazzana-Calvo, M. Gene therapy of severe combined immunodeficiencies: from mice to humans. Neth. J. Med. 60, 299-301 (2002)

13. Aiuti, A *et al*. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science. 296, 2410-2413 (2002)

14. Hacein-Bey-Abina, S. *et al*. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. 118, 3132-3142 (2008)

15. Wu, X., Li, Y., Crise, B. & Burgess, S.M. Transcription start regions in the human genome are favored targets for MLV integration. Science. 300, 1749-1751 (2003)

16. Mitchell, R.S. *et al*. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. 2, E234 (2004)

17. Lewinski, M.K. *et al*. Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS. Pathog. 2, e60 (2006)

18. Bushman, F. *et al*. Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. 3, 848-858 (2005)

19. Zufferey, R. *et al*. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J. Virol. 72, 9873-9880 (1998)

20. Miyoshi, H., Blömer, U., Takahashi, M., Gage, F.H. & Verma I.M. Development of a self-inactivating lentivirus vector. J. Virol. 72, 8150-8157 (1998)

21. Aiuti, A. *et al*. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. Science. 341, 1233151 (2013)

22. Biffi, A. *et al*. Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science. 341, 1233158 (2013)

23. Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. Annu. Rev. Genet. 40, 363-383. (2006)

24. Itzhaki, J.E. & Porter, A.C. Targeted disruption of a human interferon-inducible gene detected by secretion of human growth hormone. Nucleic Acids Res. 19, 3835-3842 (1991)

25. Brown, J.P., Wei, W. & Sedivy, J.M. Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. Science. 277, 831-834 (1997)

26. Jasin, M. & Berg, P. Homologous integration in mammalian cells without target gene selection. Genes Dev. 2, 1353-1363. (1988)

27. Lukacsovich T, Yang D, Waldman AS. Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. Nucleic Acids Res. 1994;22:5649-5657

28. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell Biol. 14, 8096-8106. (1994)

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

30. 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)

31. DeKelver, R.C. *et al*. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. Genome Res. 20, 1133-1142. (2010)

32. Lombardo, A. *et al*. Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods. 21, 861-869 (2011)

33. McConnell Smith, A. *et al*. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proc. Natl. Acad. Sci. USA. 106, 5099-5104. (2009)

34. Arnould, S. *et al*. The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Eng. Des. Sel. 24, 27-31. (2011)

35. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., Gregory, P.D. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 11, 636-646 (2010)

36. Carroll, D. Genome engineering with zinc-finger nucleases. Genetics. 188, 773-782. (2011)

37. Radecke, S., Radecke, F., Cathomen, T. & Schwarz, K. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. Mol. Ther. 18, 743-753. (2010)

38. Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. & Wolfe, S.A. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. Nucleic Acids Res. 39, 381-392. (2011)

39. Cradick, T.J., Ambrosini, G., Iseli, C., Bucher, P. & McCaffrey, A.P. ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. BMC Bioinformatics 13, 152 (2011)

40. Sander, J.D. *et al*. In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. Nucleic Acids Res. 41, e181. (2013)

41. Bhakta, M.S. *et al*. Highly active zinc-finger nucleases by extended modular assembly. Genome Res. 23, 530-538. (2013)

42. Orlando, S. *et al*. Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic Acids Res. 38, e152. (2010)

43. Doyon, Y. *et al*. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods. 8, 74-79. (2011)

44. Cristea, S. *et al*. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. Biotechnol. Bioeng. 110, 871-880. (2013)

45. Lillico, S.G. *et al*. Live pigs produced from genome edited zygotes. Sci. Rep. 10, 2847. (2013)

46. Perez, E.E. *et al*. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. 26, 808-816. (2008)

47. Holt, N. *et al*. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. Nat. Biotechnol. 28, 839-847. (2010)

48. Boch, J. *et al*. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 326, 1509-1512. (2009)

49. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. Science. 326, 1501. (2009)

50. Deng, D. *et al*. Structural basis for sequence-specific recognition of DNA by TAL effectors. Science. 335, 720-723 (2012)

51. Miller, J.C. *et al*. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29, 143-148. (2011)

52. Mussolino, C. *et al*. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 39, 9283-9293. (2011)

53. Mussolino, C. *et al*. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. Nucleic Acids Res. 42, 6762-6773. (2014)

54. Cermak, T. *et al*. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 39, e82. (2011)

55. Morbitzer, R., Elsaesser, J., Hausner, J & Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res. 39, 5790-5799. (2011)

56. Reyon, D. *et al*. FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. 30, 460-465. (2012)

57. Sanjana, N.E. *et al*. A transcription activator-like effector toolbox for genome engineering. Nat. Protoc. 7, 171-192. (2012)

58. Jinek, M. *et al*. RNA-programmed genome editing in human cells. Elife. 29, e00471. (2013)

59. Fu, Y. *et al*. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822-826. (2013)

60. Hsu, P.D. *et al*. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827-832. (2013)

61. Pattanayak, V. *et al*. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31, 839-843. (2013)

62. Mali, P. *et al*. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833-838. (2013)

63. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 32, 279-284. (2014)

64. Kuzminov, A. Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc. Natl. Acad. Sci. USA. 98, 8241-8246. (2001)

65. Tsai, Q.S. *et al*. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. 32, 569–576 (2014)

66. John P Guilinger, David B Thompson & David R Liu. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 32, 577–582 (2014)

67. Sun, N., Liang, J., Abil, Z. & Zhao, H. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. Mol. Biosyst. 8, 1255-1263. (2012)

68. Watanabe, M. *et al*. Knockout of exogenous EGFP gene in porcine somatic cells using zinc-finger nucleases. Biochem. Biophys. Res. Commun. 402, 14-18. (2010)

69. Tong, C. *et al*. Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs. J. Genet. Genomics. 39, 275-280. (2012)

70. Gaj, T., Guo, J., Kato, Y., Sirk, S.J. & Barbas, C.F. 3rd. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat. Methods. 9, 805-807. (2012)

71. Liu, J., Gaj, T., Patterson, J.T., Sirk, S.J. & Barbas, C.F. 3rd. Cell-penetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. PLoS One. 9, e85755. (2014)

72. Riick, de J. *et al*. Bromodomain and extra-terminal (BET) proteins target Moloney murine leukemia virus integration to transcription start sites. Retrovirology. 10(Suppl 1), O20. (2013)

73. Schröder, A.R. *et al*. HIV-1 integration in the human genome favors active genes and local hotspots. Cell. 110, 521-529. (2002)

74. Ranzani, M. *et al*. Lentiviral vector-based insertional mutagenesis identifies genes associated with liver cancer. Nat. Methods. 10, 155-161. (2013)

75. Kantor, B. *et al*. Epigenetic activation of HIV-1 genomes by gut-associated short chain fatty acids and its implications for HIV infection. Proc. Natl. Acad. Sci. USA. 106, 18786–18791 (2009)

76. Pelascini, L.P., Janssen, J.M. & Gonçalves, M.A. Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. Hum. Gene. Ther. 24, 78-96. (2013)

77. Pelascini, L.P. *et al.* Histone deacetylase inhibition rescues gene knockout levels achieved with integrase-defective lentiviral vectors encoding zinc-finger nucleases. Hum. Gene Ther. Methods. 24, 399-411. (2013)

78. Goodrich, D.W. & Duesberg, P.H. Retroviral recombination during reverse transcription. Proc. Natl. Acad. Sci. USA. 87, 2052-2056. (1990)

79. Hu, W.S. & Temin, H.M. Retroviral recombination and reverse transcription. Science. 250, 1227-1233. (1990)

80. Holkers, M. *et al*. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Res. 41, e63. (2013)

81. Holkers, M., Cathomen, T. & Gonçalves, M.A. Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells. Methods. 2023, 61-69 (2014)

82. Yang, L. *et al*. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 41, 9049-9061. (2013)

83. Cai, Y., Bak, R.O. & Mikkelsen, J.G. Targeted genome editing by lentiviral protein transduction of zinc-finger and TAL-effector nucleases. Elife. 3:e01911 (2014)

84. Ellis, B.l., Hirsch, M.L., Porter, S.N., Samulski, R.J. & Porteus, M.H. Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by Food en Drug Administration-approved drugs. Gene Ther. 20, 35-42 (2013)

85. Gonçalves, M.A. Adeno-associated virus: from defective virus to effective vector. Virol. J. 2, 43. (2005)

86. Ferrari, F.K., Samulski, T., Shenk, T. & Samulski, R.J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J. Virol. 70, 3227-3234. (1996)

87. Fisher, K.J. *et al*. Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J. Virol. 70, 520-532. (1996)

88. McCarty, D.M., Monahan, P.E. & Samulski, R.J. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther. 8, 1248-1254. (2001)

89. De Jong, J.C. *et al*. Adenoviruses from human immunodeficiency virus-infected individuals, including two strains that represent new candidate serotypes Ad50 and Ad51 of species B1 and D, respectively. J. Clin. Microbiol. 37, 3940-3945. (1999)

90. Horwitz, M.S. Adenovirus immunoregulatory genes and their cellular targets. Virology. 279, 1-8. (2001)

91. Graham, F.L., Smiley, J., Russell, W.C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. 36:59–74 (1977)

92. Fallaux, F.J. *et al*. Characterization of 911: A New Helper Cell Line for the Titration and Propagation of Early Region 1-Deleted Adenoviral Vectors. Hum. Gene Ther. 7:215–222 (1996)

93. Fallaux, F.J. *et al*. New Helper Cells and Matched Early Region 1-Deleted Adenovirus Vectors Prevent Generation of Replication-Competent Adenoviruses. Hum. Gene. Ther. 9:1909–1917 (1998)

94. He, T.C., Zhou, S., da Costa, L.T., Yu, J., Kinzler, K.W. & Vogelstein, B. A simplified system for generating recombinant adenoviruses. Proc. Natl. Acad. Sci. USA. 95:2509-2514 (1998)

95. Janssen, J.M., Liu, J., Skokan, J., Gonçalves, M.A. & de Vries, A.A. Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. J. Gene Med. 15, 1-11 (2013)

96. Havenga, M.J. *et al*. Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: High yield and preserved bioactivity. Biotechnol. Bioeng. 100:273–283 (2008)

97. Gonçalves, M.A. *et al*. Efficient Generation and Amplification of High-Capacity Adeno-Associated Virus/Adenovirus Hybrid Vectors. J. Virol. 76:10734-10744 (2002)

98. Yuan, J. *et al*. Zinc-finger nuclease editing of human cxcr4 promotes HIV-1 CD4(+) T cell resistance and enrichment. Mol. Ther. 20:849-859 (2012)

99. Holkers, M. *et al*. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Sci. Rep. 29, 5105 (2014)

100. Disterer, P., Papaioannou, I., Evans, V.C., Simons, J.P. & Owen, J.S. Oligonucleotide-mediated gene editing is underestimated in cells expressing mutated green fluorescent protein and is positively associated with target protein expression. J. Gene. Med. 14, 109-119. (2012)

101. Papaioannou, I., Simons, J.P. & Owen, J.S. Oligonucleotide-directed gene-editing technology: mechanisms and future prospects. Expert Opin. Biol. Ther. 12, 329-342. (2012)

102. Lin, F.L., Sperle, K. & Sternberg, N. Repair of double-stranded DNA breaks by homologous DNA fragments during transfer of DNA into mouse L cells. Mol. Cell. Biol. 10:113-119 (1990)

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

104. Lombardo, A. *et al*. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat. Biotechnol. 25:1298–1306 (2007)

105. Moiani, A. *et al*. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. J. Clin. Invest. 122:1653-1666 (2012)

106. Delviks, K.A. & Pathak, V.K. Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching. J. Virol. 73:7923–7932 (1999)

107. Khan, I.F., Hirata, R.K. & Russell. D.W. AAV-mediated gene targeting methods for human cells. Nat. Protoc. 6:482-501 (2011)

108. Bordignon, C. *et al*. Transfer of the HSV-tk gene into donor peripheral blood lymphocytes for in vivo modulation of donor anti-tumor immunity after allogeneic bone marrow transplantation. Hum. Gene Ther. 6:813-819 (1995)

109. Eissenberg, L.G., Rettig, M., Dehdashti, F., Piwnica-Worms, D. & DiPersio, J.F. Suicide genes: monitoring cells in patients with a safety switch. Front. Pharmacol. 6:241 (2014)

110. Suzuki, K. *et al*. Highly efficient transient gene expression and gene targeting in primate embryonic stem cells with helper-dependent adenoviral vectors. Proc. Natl. Acad. Sci. USA 16:13781-13786 (2008)

111. Karreman, S., Hauser, H. & Karreman, C. On the use of double FLP recognition targets (FRTs) in the LTR of retroviruses for the construction of high producer cell lines. Nucl. Acid. Res. 24:1616-1624 (1996)

112. Kuehle, J. *et al*. Modified lentiviral LTRs allow Flp recombinase-mediated cassette exchange and in vivo tracing of "factor-free" induced pluripotent stem cells. Mol. Ther. 22:919-928 (2014)

113. Chang, C.W. *et al*. Polycistronic lentiviral vector for "hit and run" reprogramming of adult skin fibroblasts to induced pluripotent stem cells. Stem Cells. 27:1042-1049 (2009)

114. Sauer B. Identification of cryptic lox sites in the yeast genome by selection for Cre-mediated chromosome translocations that confer multiple drug resistance. J. Mol. Biol. 20:911-928 (1992)

115. Nagy A. Cre recombinase: the universal reagent for genome tailoring. Genesis 26:99-109 (2000)

116. Liu, G.H. *et al*. Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. Cell Stem Cell. 8:688-694 (2011)

117. Mitani K. Gene targeting in human-induced pluripotent stem cells with adenoviral vectors. Methods Mol. Biol. 1114:163-167 (2014)

1

2

Non-spaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells

Maarten Holkers, Antoine A.F. de Vries, and Manuel A.F.V. Gonçalves

Department of Molecular Cell Biology, Leiden University Medical Center,
Einthovenweg 20, 2333 ZC Leiden, the Netherlands

Non-spaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells

Maarten Holkers, Antoine A.F. de Vries, and Manuel A.F.V. Gonçalves

Department of Molecular Cell Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands

ABSTRACT

DNA repeats constitute potential sites for the nucleation of secondary structures such as hairpins and cruciforms. Studies performed mostly in bacteria and yeast showed that these non-canonical DNA structures are breakage-prone, making them candidate targets for cellular DNA repair pathways. Possible culprits for fragility at repetitive DNA sequences include replication and transcription as well as the action of structure-specific nucleases. Despite their patent biological relevance, the parameters governing DNA repeat-associated chromosomal transactions remain ill-defined. Here, we established an episomal recombination system based on donor and acceptor complementary DNA templates to investigate the role of direct and inverted DNA repeats in homologous recombination in mammalian cells. This system allowed us also to ascertain in a stringent manner the impact of repetitive sequence replication on homology-directed gene repair. We found that non-spaced DNA repeats can, *per se*, engage the homologous recombination pathway of the cell and that this process is primarily dependent on their spacing and relative arrangement (i.e. parallel or anti-parallel) rather than on their sequence. Indeed, our data demonstrate that contrary to direct and to spaced inverted repeats, non-spaced inverted repeats are intrinsically recombinogenic motifs in mammalian cells lending experimental support to their role in genome dynamics in higher eukaryotes.

INTRODUCTION

The genomes of prokaryotes and eukaryotes harbor numerous and diverse types of repetitive DNA sequences many of which have been associated with genome evolution, regulation of gene expression and chromosomal rearrangements underlying a number of inherited disorders and certain translocation-bearing tumors. These motifs include single direct and inverted DNA repeats with or without internal spacers as well as high-copy-number tandem tracts (1-4). Accumulating evidence indicates that DNA repeats can adopt different non-canonical (i.e. non-B) DNA conformations depending on a number of intrinsic parameters. These include the nucleotide composition, the length and the relative orientation of the constituent DNA units as well as their spacing and extent of sequence identity. Extrinsic factors such as the torsional strain associated with DNA metabolic processes, chromatinization and transcription are also thought to influence the likelihood that DNA repeats acquire higher-order conformations. DNA conformers have been implicated in both physiological and pathological processes including the regulation of DNA replication and expression, oncogenic chromosomal rearrangements and gene amplification (1-7). Related to this, palindromes (i.e. uninterrupted or non-spaced inverted DNA repeats) and inverted DNA repeats with relatively short central spacers, can, via local negative superhelical stress and ensuing intrastrand hybridization and branch migration, extrude into four-way Holliday junction-like DNA structures or cruciforms. Inverted DNA repeats in single-stranded form may also originate stem-loops or hairpins via intrastrand annealing. This may, for instance, occur when the unwinding of double-helical DNA during replication creates a lagging strand template.

The rearrangement of chromosomal DNA carrying non-canonical structures are likely preceded by and dependent on phosphodiester bond cleavage presumably via their resolution and processing by structure-specific nucleases. This might occur in concert with DNA replication-associated phenomena such as replisome stalling or slippage. In *Escherichia coli*, physical evidence was recently obtained for the emergence of double-stranded DNA breaks (DSBs) at a 246-base-pair (bp) palindrome via the combined effects of DNA replication and cleavage by the Mre11/Rad50 homolog SbcCD (8). Interestingly, studies carried out in a yeast model system revealed that inverted repeats of a 320-bp retrotransposon-derived human *Alu* sequence inserted into a *LYS2* reporter allele were a target for the Mre11/Rad50/Nbs1 complex suggesting evolutionary conservation of DNA structure-processing biochemical pathways (9). Moreover, in these prokaryotic and lower eukaryotic model systems, reporter gene expression rescue assays showed that long (i.e. >150 bp) inverted repeat-associated DNA breaks could engage the error-free homologous recombination (HR) pathway (3). Notwithstanding steady progress in this field, many questions remain with respect to the relationships between specific parameters of repetitive DNA motifs, putative

ensuing higher-order DNA conformations and the recruitment of cellular pathways that regulate genetic recombination. This knowledge gap is particularly acute in cells of higher eukaryotes (2,3). In addition, hitherto, the vast majority of studies on the biological activity and fate of repetitive DNA *in vivo* focused on endogenous or exogenous test sequences embedded within the chromosomal DNA of dividing cells. With this type of experimental setups is difficult to assess a possible contribution of template DNA replication to repeat-associated phenomena.

Here, we developed and deployed an extrachromosomal recombination system to specifically address the role of single DNA repeats of different sequence, arrangement (i.e. parallel or anti-parallel) and spacing in HR-mediated DNA repair in mammalian cells. Furthermore, the introduction of a eukaryotic origin of replication into the repetitive DNA-containing episomes allowed us to also investigate the impact of target template DNA replication on the recombinogenic potential of the various motifs. We demonstrate that simple palindromes and composite inverted DNA repeats, but not direct or spaced inverted DNA repeats, serve as targets for the error-free HR repair pathway in mammalian cells and that this process is independent of ongoing DNA repeat-bearing molecule replication.

MATERIALS AND METHODS

Cells
HeLa cells (American Type Culture Collection [ATCC]), human embryonic kidney (HEK) 293T cells (ATCC) and 911 cells (10) were cultured in Dulbecco's modified Eagle's medium (D-MEM; Invitrogen) supplemented with 5% fetal bovine serum (FBS; Invitrogen). PER.tTA. Cre76 cells (11) and COS-7 cells (ATCC) were propagated in D-MEM supplemented with 10% FBS. All cells were cultured at 37°C in an atmosphere of 10% CO_2 in humidified air.
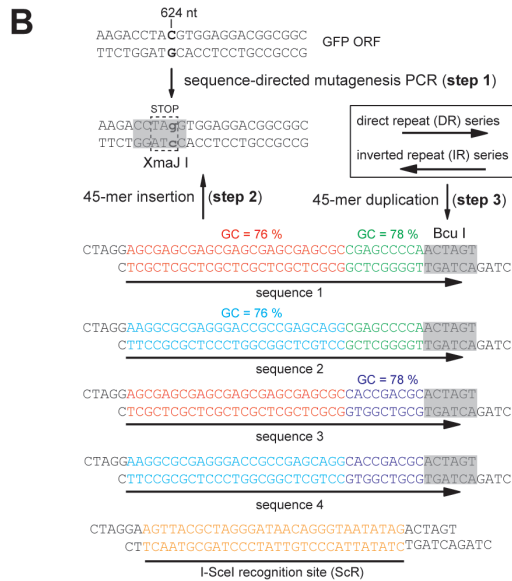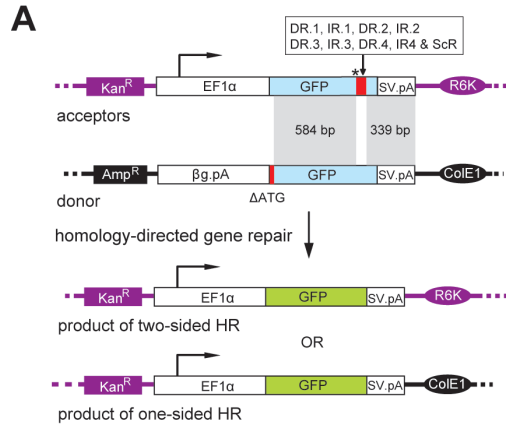
Recombinant DNA. Plasmid pA1.GFP.A2 has been described previously (GenBank accession number: GQ380658 [12]). An XmaJI recognition sequence was introduced in pA1.GFP. A2 at nucleotide positions 620 through 625 of the humanized *Renilla reniformis* green fluorescent protein (*GFP*) open reading frame (ORF) by PCR site-directed mutagenesis to generate the acceptor plasmid pR6K.GFP.STOP. Moreover, pR6K.GFP.STOP has the *GFP* ORF disrupted by an amber stop codon (Figure 1). The nucleotide sequences of the sense and antisense primers used for introducing the mutation that created the XmaJI site were 5'-GAAGACCTA*G*GTGGAGGAC-3' and 5'-GTCCTCCAC*C*TAGGTCTTC-3', respectively (point mutation is underlined) and the PCR was carried out with Phusion High-Fidelity DNA polymerase (Finnzymes) according to the instructions provided by the manufacturer. Oligodeoxyribonucleotides used for the introduction of DNA sequences into the XmaJI site of

pR6K.GFP.STOP were 5'-CTAGGAGCGAGCGAGCGAGCGAGCGAGCGCCGAGCCCCAACTAGT-3' and 5'-CTAGACTAGTTGGGGCTCGGCGCTCGCTCGCTCGCTCGCTCGCTC-3' (DR/IR.1), 5'-CTAGGAAGGCGCGAGGGACCGCCGAGCAGGCGAGCCCCAACTAGT-3' and 5'-CTAGACTAGTTGGGGCTCGCCTGCTCGGCGGTCCCTCGCGCCTTC-3' (DR/IR.2), 5'-CTAGAGACGACGCAGCGAGCGAGCGAGCGCCACCGACGCACTAGT-3' and 5'-CTAGACTAGTGCGTCGGTGGCGCTCGCTCGCTCGCTGCGTCGTCT-3' (DR/IR.3) and 5'-CTAGGAAGGCGCGAGGGAGGGACCGCCGAGCAGGCACCGACGCACTAGT-3' and 5'-CTAGACTAGTGCGTCGGTGCCTGCTCGGCGGTCCCTCGCGCCTTC-3' (DR/IR.4). Insertion of a recognition sequence for the meganuclease I-SceI into the XmaJI site of pR6K.GFP.STOP was accomplished using the oligodeoxyribonucleotides 5'-CTAGGAAGTTACGCTAGGGATAACAGGGTAATATAGACTAGT-3' and 5'-CTAGACTAGTCTATATTACCCTGTTATCCCTAGCGTAACTTC-3'. To generate pR6K.GFP.STOP derivatives containing DNA repeats in a head-to-tail (direct repeat) or tail-to-tail (inverted repeat) configuration, the constructs carrying single copies of the oligodeoxyribonucleotide pairs corresponding to DR/IR.1, DR/IR.2, DR/IR.3 and DR/IR.4 were linearized with BcuI and subsequently subjected to a second round of oligodeoxyribonucleotide cloning. Restriction fragment size analysis was used to distinguish between recombinant plasmids carrying a direct or an inverted repeat of each oligodeoxyribonucleotide pair. To disrupt the palindrome in the IR.1-containing pR6K.GFP.STOP derivative acceptor[IR.1] (Figure 1B and Figure 1C) at the center of symmetry, the plasmid was digested with BcuI and its backbone was combined with the oligodeoxyribonucleotide pair containing the I-SceI recognition sequence (ScR). The resulting construct was designated acceptor[spIR.1]. To create an acceptor plasmid, in which the *GFP* ORF is interrupted by the composite adeno-associated virus type 2 (AAV) inverted terminal repeat (ITR), AAV vector shuttle plasmid pDD2 (13) was digested with PvuII and BspLI. The resulting 127-bp AAV ITR-specific DNA fragment was inserted into the XmaJI site of pR6K.GFP.STOP following its blunt-ending with the Klenow fragment of *Escherichia coli* DNA polymerase I (Klenow, Fermentas) to produce acceptor[ITR]. Plasmid pUC.hrGFPI.SV40pA was prepared by introducing the *GFP* ORF and the downstream bidirectional simian virus 40 (SV40) polyadenylation signal (pA) derived from pA1.GFP.A2 into pUC19 using BamHI and XbaI. Next, the pA of the *rabbit β-globin* (*βG*) gene was inserted immediately upstream of the GFP-coding sequence in pUC.hrGFPI.SV40pA to inhibit possible spurious transcription of the *GFP* ORF due to the presence of cryptic promoters in the plasmid backbone. To this end, pAAV.hEF1a.DsRedT4.rbGpA, an AAV vector shuttle construct containing a *DsRed.T4* (14) expression unit controlled by the *human eukaryotic translation elongation factor 1 alpha 1* (*EF1α*) gene promoter and *βG* pA, was incubated with NotI and SmiI, the digestion products were blunt-ended using Klenow and the 587-bp *βG* pA-containing DNA fragment was purified from agarose gel. Subsequently, this fragment was inserted in the proper orientation into pUC.hrGFPI.SV40pA between the Klenow-blunted XbaI and HindIII sites

to produce pUC.donor.rbGpA.hrGFPI.SV40pA. Deletion of the *GFP* start codon from pUC.donor.rbGpA.hrGFPI.SV40pA was achieved by digesting the plasmid with SalI and SdaI, filling in the 3' recessed ends with Klenow and self-ligation of the plasmid backbone to create the donor template pUC.donor.GFP.ΔATG (i.e. GFP^{ΔATG}; GenBank accession number: JF714898). Construct pUC.donor.rbGpA.DsRed.T4.SV40pA was made by replacing the *GFP* ORF in pUC.donor.rbGpA.hrGFPI.SV40pA by the coding sequence of the red fluorescent protein (RFP) DsRed.T4. The DsRed.T4-coding sequence was excised from pAAV.hEF1a.DsRedT4.rbGpA using XbaI and NotI and subsequently combined with the 3.5-kb XbaI×NotI fragment of pUC.donor.rbGpA.hrGFPI.SV40pA. Disruption of the *RFP* ORF in pUC.donor.rbGpA.DsRed.T4.SV40pA was accomplished by linearization with NcoI followed by Klenow treatment and self-ligation resulting in non-homologous donor plasmid pUC.donor.RFP.ΔATG (i.e. RFP^{ΔATG}; GenBank accession number: JF714899). The *I-SceI* expression construct pCAG.I-SceI (GenBank accession number: JF714900) was made by inserting the 3.1-kb SalI×PstI fragment of pCbASce (15) into the polylinker of pUC19 after its digestion with SalI and PstI. The expression plasmid pCAG.I-SceI(Δ112-246) encoding a non-functional version of I-SceI was generated by digesting pCAG.I-SceI with BstBI followed by self-ligation of the resulting vector backbone. All DNA preparations were made by using the JetStar 2.0 DNA isolation system (Genomed).

Figure 1. Episomal recombination system to study the effects of repetitive DNA sequences on homology-directed gene repair in mammalian cells. (**A**) The bipartite system consists of acceptor and donor plasmids with differently disrupted *GFP* ORFs. In the acceptor plasmids, the *GFP* ORF is interrupted by an amber stop codon (asterisk) plus a test DNA sequence of choice (large red bar), whereas in the donor plasmid it is rendered non-functional by the deletion of its first 38 nts (short red bar). The interrupted *GFP* ORF in the various acceptor plasmids is framed by a constitutively active (broken arrow) human *eukaryotic translation elongation factor 1 alpha 1* gene (EF1α) promoter and the SV40 pA (SV.pA), whereas that in the donor plasmid is preceded by the rabbit *β-globin* pA (βG.pA) and followed by the SV40 pA. The prokaryotic origins of replication R6K and ColE1 as well as the antibiotic resistance genes *aminoglycoside 3'-phosphotransferase* (Kan^R) and *β-lactamase* (Amp^R) present in the acceptor and donor plasmid backbones, respectively, are also indicated. Once introduced into cells, donor and acceptor plasmids are candidate substrates for HR by virtue of the shared 339- and/or 584-bp DNA sequences. Reciprocal exchange of genetic information between acceptor and donor templates via cross-overs within these homologous regions is expected to give rise to transcription units with restored ORFs directing the synthesis of full-length GFP. (**B**) Three-step strategy to generate the various acceptor plasmids with test sequences 1 through 4 arranged in a direct or inverted repeat orientation (DR and IR series, respectively). Step 1: PCR-based site-directed mutagenesis of the C residue at position 624 of the *GFP* ORF into a G generates a premature stop codon and a XmaJI recognition site. Step 2: Insertion, at the XmaJI site, of test sequences 1, 2, 3 or 4, which all contain a BcuI recognition sequence at their 3' end. Step 3: Molecular clones with the BcuI sites in a position distal to the premature stop codon (orientation depicted) were used to duplicate test sequences 1, 2, 3 or 4. Acceptor plasmids with the meganuclease I-SceI recognition site or containing a single copy of target sequence 1 served as controls. (**C**) Schematic representation of directed and inverted repeats of DNA sequences 1 through 4, the I-SceI recognition site and a single copy of test sequence 1. The propensity of each DNA sequence to transit from lineform to cruciform by intrastrand Watson and Crick base pairing was estimated by calculating the Gibbs free energy (ΔG) in the presence of 150 mM NaCl or of 1M NaCl. ▶

To provide acceptor[ScR], acceptor[DR.1], acceptor[IR.1] and acceptor[spIR.1] with an SV40 origin of replication (ori), they were linearized with NdeI and blunt-ended using Klenow. Next, these linear DNA molecules were ligated to the SV40 ori-containing 323-bp PvuII×Eco147I fragment of pGL4.22 (GenBank accession number: DQ188842), yielding acceptor[ScR.ORI], acceptor[DR.1.ORI], acceptor[IR.1.ORI] and acceptor[spIR.1.ORI]. All oligodeoxyribonucleotides were supplied by Eurofins MWG Operon, while the restriction and DNA modifying enzymes were from Fermentas.

Gibbs free energy calculations

The Gibbs free entry of the most stable secondary structure that could be folded by each of the DNA segments inserted into the XmaJI site of pR6K.GFP.STOP was calculated with the aid of the software program Mfold 3.2 (16) using energy rules for DNA (17) at http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form.

Extrachromosomal DNA extraction. Extrachromosomal DNA was extracted from the transfected cells essentially as described before (18) Briefly, at 72 hours post-transfection, cells were scraped from the surface of a 2-cm² well with the plunger of a 1-ml Luer-Lok disposable syringe (BD Biosciences). The cell suspension was collected in a 15-ml screwcap tube with a conical bottom (Greiner Bio-One) and centrifuged for 5 minutes at 1,500×g. The supernatant was aspirated and the cells were washed once with 5 ml phosphate-buffered saline. After another round of centrifugation, the cell pellet was resuspended in 180 μl of solution I (10 mM Tris-HCl at pH 8.0; 10 mM EDTA; 100 μg/ml proteinase K [Fermentas]) and transferred to a 1.5-ml microtube (Eppendorf). Next, 180 μl of solution II (10 mM Tris-HCl at pH 8.0; 10 mM EDTA; 1.2% sodium dodecyl sulfate) was added. The microtube was inverted thrice to mix its content and incubated for 30 minutes at 37°C. Next, the sample was mixed with 90 μl of 5 M NaCl and stored overnight at 4°C. The following day, the chromosomal DNA was pelleted by centrifugation for 60 minutes at 16,100×g and the supernatant was removed to a new microtube. Subsequently, the supernatant was extracted twice with buffer-saturated phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform and the episomal DNA was precipitated by addition of 2.5 volumes of ethanol and 0.5 volumes of 7.5 M ammonium acetate at pH 5.5. After washing with 70% ethanol, the DNA pellet was dried and dissolved in 100 μl of TE⁺ buffer (10 mM Tris-HCl at pH 8.0; 1 mM EDTA; 100 μg/ml RNase A [Fermentas]). The purified DNA was used for PCR and Southern blot analyses.

Southern blot analysis. One fifth (i.e. 20 μl) of the extracted extrachromosomal DNA was incubated with XbaI and DpnI. XbaI releases the *GFP* ORF plus downstream SV40 pA from the acceptor plasmids for easy screening, whereas DpnI selectively digests the prokaryotic input DNA. The resulting DNA fragments were separated in a 0.7% agarose gel in 1× TAE buffer.

Next, the DNA was transferred by capillary action to an Amersham Hybond-XL membrane (GE Healthcare) using a standard Southern blot technique. The 744-bp *GFP*-specific probe was obtained by digestion of plasmid pA1.GFP.A2 with XbaI and XhoI (both from Fermentas) followed by preparative agarose gel electrophoresis. The DNA probe was labeled with EasyTide (α-^{32}P) dCTP (3000 Ci/mmol; Perkin Elmer) using the DecaLabel DNA Labeling Kit (Fermentas). Prior to their application in hybridization experiments, the radiolabeled DNA fragments were separated from unincorporated dNTPs through size exclusion chromatography using Sephadex-50 (GE Healthcare) columns. A Storm 820 PhosphorImager (Amersham Biosciences) was used for the detection of labeled DNA. Images were acquired using the Storm scanner control 5.03 software and processed using ImageQuant Tools 3.0 software (both from Amersham Biosciences).

In vivo assay to detect processing of inverted DNA repeats

Eighty-thousand HeLa cells were transfected essentially as described under "DNA transfections" with 266 ng of acceptor[DR.1], acceptor[IR.1] or acceptor[ScR] each mixed with 133 ng of pCAG.I-SceI or of pCAG.I-SceI(Δ112-246). Extrachromosomal DNA was isolated 72 hours post-transfection as described elsewhere in this section and PCR was performed on 4 µl of DNA using 0.4 µM of primers 1 (5'-ATGGTGAGCAAGCAGATCCTGAAG-'3) and 2 (5'-CCGAGAAGGAAGTGCTCC-3'), 0.4 mM of each dNTP (New England Biolabs), 1×GoTaq reaction buffer, 1 mM MgCl$_2$ and 2.5 U of GoTaq DNA polymerase (all from Promega) in a final volume of 50 µl. The PCR cycles were performed in a DNA Engine Tetrad 2 Peltier Thermal Cycler (Bio-Rad) using the following cycling conditions. A first denaturating step at 95°C for 5 min, followed by 30 cycles of 60 sec at 95°C, 60 sec at 64°C and 120 sec at 72°C. Reactions were terminated by a final extension period of 5 min at 72°C. The synthesized DNA was purified using SureClean (Bioline) and dissolved in 30 µl of 10 mM Tris-Cl pH 8.0. The resulting PCR products were treated with DpnI alone or with DpnI plus I-SceI, XmaJI or BcuI and, subsequently, subjected to agarose gel electrophoresis. The inclusion of DpnI served to remove possible residual input prokaryotic plasmid DNA prior to Sourthern blot analysis. The procedures for Southern blotting and for DNA probe radiolabeling are described elsewhere in this section. The probe used is complementary to the *hrGFP* ORF and contiguous SV40 polyadenylation signal sequences. Undigested PCR products were cloned in the pCR4-TOPO cloning vector (invitrogen) using GT115 chemically-competent *E. coli* cells. Individual molecular clones corresponding to independent DNA processing events were sequenced using primer 1 or primer 6 (5'-CAGCTTCGAGGTGGTG-3').

Statistical analysis

Statistical parameters were computed using Graph Pad Prism 4.03. Student's t-test was applied to compare data sets with $P < 0.05$ being considered significant.

RESULTS

Design and construction of the episomal HR assay system

We established an extrachromosomal assay system to study the role of DNA repeats in homology-directed gene repair *in vivo*. This system is based on pairs of recombination substrates consisting of donor and acceptor (or target) DNA molecules containing *GFP* reporter genes that are rendered defective by different means (Figure 1A). The transcription unit in the donor plasmid was made non-functional by removing the first 38 nucleotides from the *GFP* ORF and lacks eukaryotic promoter/enhancer elements whereas the *GFP* ORF in the various acceptor constructs was disrupted by an internal stop codon preceding the test DNA sequences depicted in Figure 1B and Figure 1C. The donor and acceptor DNA templates share two regions of perfect sequence identity (Figure 1A, marked in grey), which in the acceptor plasmids are separated from each other by the test sequences. HR-dependent reciprocal exchange of genetic information through the common DNA segments of 584 and 339 bp or via a single cross-over within the upstream, 584-bp, arm of homology are both expected to generate fully functional *GFP* transcription units. Thus, the capacity of different test DNA sequences to elicit homology-directed gene repair can be readily assessed by the analysis of *GFP* expression through direct fluorescence microscopy or flow cytometry.

The generation of site-specific DSBs by the *Saccharomyces cerevisiae* mitochondrial group I intron-encoded homing endonuclease I-SceI at its cognate 18-bp recognition sequence is a well-established method to trigger in a predictable manner DNA repair pathways in prokaryotic and eukaryotic systems (19) Therefore, in our experimental system, we used as positive control for HR-dependent rescue of *GFP* expression, cells exposed to a donor plasmid, an acceptor template containing the I-SceI recognition sequence (ScR) and an *I-SceI* expression construct. Cells that only received the first two plasmids served as negative control in the HR assay.

Using PCR-based site-directed mutagenesis, an XmaJI cleavage site was introduced into the GFP-coding sequence of the starting construct that served as basis to generate the acceptor plasmid panel (Figure 1B). This maneuver resulted also in the disruption of the GFP *ORF* by a stop codon contained within the XmaJI recognition sequence. Next, DNA fragments composed of hybridized synthetic oligodeoxyribonucleotides with XmaJI-compatible cohesive termini encompassing the test sequences 1 through 4 (45-mers) or the recognition sequence for the I-SceI meganuclease (42-mers; Figure 1B, orange) were individually inserted at the PCR-created XmaJI site (Figure 1B). The four test sequences consisted of two different DNA segments of 25 bp (Figure 1B, red and cyan) extended at the 3' end with two different 9-bp DNA sequences (Figure 1B, green and violet) and a BcuI recognition sequence (Figure 1B, grey). Subsequently, another copy of each of the four test sequences was added at the BcuI site leading to acceptor plasmids in which the *GFP* ORF is

interrupted by the direct repeats DR.1, DR.2, DR.3 or DR.4 or the inverted repeats IR.1, IR.2, IR.3 or IR.4 (Figure 1C). Due to their identical GC content, the four different test sequences were predicted to possess very similar and low folding free energies when in a tandem antiparallel orientation while those corresponding to their arrangement in a direct repeat configuration were calculated to be rather high (Figure 1C). Thus, on theoretical grounds the inverted repeats of the four test sequences have a higher likelihood to originate secondary structures via intrastrand hybridization than their isogenic counterparts displaying a direct repeat arrangement (Figure 1C).

Experimental evidence for inverted repeat-dependent formation of DNA secondary structures in acceptor plasmids

The bacteriophage T7 endonuclease I is a commonly used tool to probe for the presence of cruciform- or Holliday-like secondary structures in DNA. This resolvase recognizes preferentially the four-way junctions characteristic of these types of DNA conformers and, in its dimeric form, introduces paired nicks close to the branch points leading to subsequent DNA cleavage. On the other hand, negative superhelical torsional stress, such as present in the supercoiled fraction of plasmid DNA (SC), is a major driving force in the nucleation and extrusion of cruciform-like structures at DNA repeats. Thus, the *E. coli* enzyme DNA topoisomerase I by catalyzing the relaxation of negatively supercoiled DNA should, to a great extent, inhibit the generation of cruciforms as scored in *in vitro* assays based on the accumulation of T7 endonuclease I-resolved plasmid DNA molecules (i.e. linear form). The results presented in Figure 2A correspond to an experiment in which the validated cruciform-forming plasmid pUC(AT) was used to confirm that the time-dependent accumulation of the linear T7 endonuclease I-derived DNA product does indeed depend on SC DNA. In fact, in the presence of the DNA topoisomerase I, virtually all the pUC(AT) plasmid transited from the SC to the OC form resulting in a concomitant stringent inhibition of cruciform resolution. Importantly, an equivalent experimental outcome was observed when the DNA substrate harboring the test IR.1 sequence was deployed, suggesting that this plasmid is prone to the acquisition of cruciform-like structures as well (Figure 2B).

Next, to investigate the ability of the different tandem arrangements of the test DNA sequences to induce the nucleation and extrusion of cruciform-like structures in plasmid substrates, the ScR-, IR.1- or DR.1-containing acceptor constructs were treated or mock-treated with T7 endonuclease I. For the sake of clarity, hereinafter these acceptor plasmids will be named after the test sequences that they contain (e.g. acceptor[ScR], acceptor[IR.1], and so forth). Because acceptor[ScR] purposely possesses a test DNA sequence with an intrinsically low intrastrand folding capacity (ΔG = -0.35 kcal/mol) it served to establish the background of the assay. To provide for a test DNA sequence known to give rise to four-way DNA structures under physiological conditions, we generated acceptor[ITR], which has a 127-bp DNA segment

derived from the AAV ITR inserted at *GFP* nucleotide position 624. This DNA segment, which contains three self-complementary regions (i.e. A/A', B/B' and C/C'), has a high propensity to fold into a T-shaped hairpin structure by intrastrand hybridization (20,21). Thus, the acceptor[ITR] plasmid was also subjected to T7 endonuclease I treatment. Furthermore, to generate internal references, each acceptor plasmid was linearized in parallel with the restriction enzyme ApaLI.

Agarose gel electrophoresis of T7 endonuclease I- or ApaLI-treated acceptor plasmids revealed that fragments with sizes consistent with DNA cleavage (i.e. linearized DNA forms) were clearly more prominent in the samples of acceptor[IR.1] and acceptor[ITR] than in those corresponding to acceptor[ScR] and acceptor[DR.1] (Figure 2C). Indeed, there was no noticeable difference in the accumulation of linear DNA between the non-repeat-containing acceptor[ScR] and the direct repeat-containing acceptor[DR.1]. As expected, agarose gel electrophoresis of acceptor plasmids not exposed to T7 endonuclease I resulted in the detection of only supercoiled and open circular (i.e. nicked) DNA topologies (Figure 2C).

Figure 2. *In vitro* assay to probe for the formation of cruciform-like secondary structures in DNA repeat-containing acceptor plasmids. (**A**) Testing the effect of the negative supercoiling-relaxing enzyme DNA topoisomerase I on the yields of T7 endonuclease I-resolved DNA (solid arrowhead) using the validated cruciform-forming plasmid pUC(AT). pUC(AT) treated or not treated with DNA topoisomerase I (+Topo and -Topo, respectively) was exposed to T7 endonuclease I for 10, 20, 30 or 60 min or underwent a 60-min mock treatment (0). Following agarose gel electrophoresis, the resulting DNA forms were visualized through ethidium bromide staining. OC and SC, open circular and supercoiled DNA forms, respectively. (**B**) Testing the impact of the negative supercoiling-relaxing enzyme DNA topoisomerase I on the yields of T7 endonuclease I-resolved DNA (solid arrowhead) using the acceptor substrate containing the test non-spaced inverted repeat sequence IR.1. Acceptor[IR.1] treated or not treated with DNA topoisomerase I (+Topo and −Topo, respectively) was exposed to T7 endonuclease I for 10, 20, 30 or 60 min or was subjected to a 60-min mock treatment (0). C and SC, open circular and supercoiled DNA forms, respectively. (**C**) Target DNA plasmids acceptor[ScR] (ScR), acceptor[IR.1] (IR.1), acceptor[DR.1] (DR.1) and acceptor[ITR] (ITR) were incubated in the presence (+) or absence (-) of T7 endonuclease I (T7) for 10 min. The resulting DNA products were resolved by agarose gel electrophoresis and stained with ethidium bromide. OC and SC, open circular and supercoiled DNA forms, respectively. L, Acceptor plasmid linearized with ApaLI. (**D**) Target DNA plasmids acceptor[IR.1], acceptor[DR.1] and acceptor[ITR] were either incubated with T7 endonuclease I for 10, 20, 30 or 60 min or underwent a 60-min mock treatment (0). The resulting DNA products were analyzed by agarose gel electrophoresis and ethidium bromide staining. L, Acceptor plasmid linearized with ApaLI. (**E**) The test plasmids harboring the non-spaced inverted repeat sequences IR.1 through IR.4 (left panels) as well as those containing their respective direct repeat counterparts (right panels) were treated with T7 endonuclease I for 20, 30 or 60 min prior and resolved through agarose gel electrophoresis. SC, supercoiled DNA; solid arrowheads point to the resolved linear molecules. (**F**) The plasmids acceptor[DR.1], acceptorsp[IR.1], acceptor[IR.1] and acceptor[ITR] were incubated with T7 endonuclease I for 10, 20, 30 and 60 min and analyzed by agarose gel electrophoresis. (**G**) *In vitro* mapping of the T7 endonuclease I cleavage site in acceptor molecules. Upper panel, diagram of the expected digestion patterns resulting from the combined activities of SalI and T7 endonuclease I or of HincII and T7 endonuclease I. The numerals correspond to the sizes (in bp) of the different DNA fragments each of which drawn in relation to the parental acceptor DNA template containing the ITR sequence embedded within the *hrGFP* ORF. Lower left and right panels, agarose gel electrophoresis of ITR-containing acceptor molecules treated only with SalI or with SalI and T7 endonuclease I or exposed to HincII or to HincII plus T7 endonuclease I, respectively. Lanes M in all the panels, GeneRuler DNA Ladder Mix molecular weight marker (Fermentas). ▶
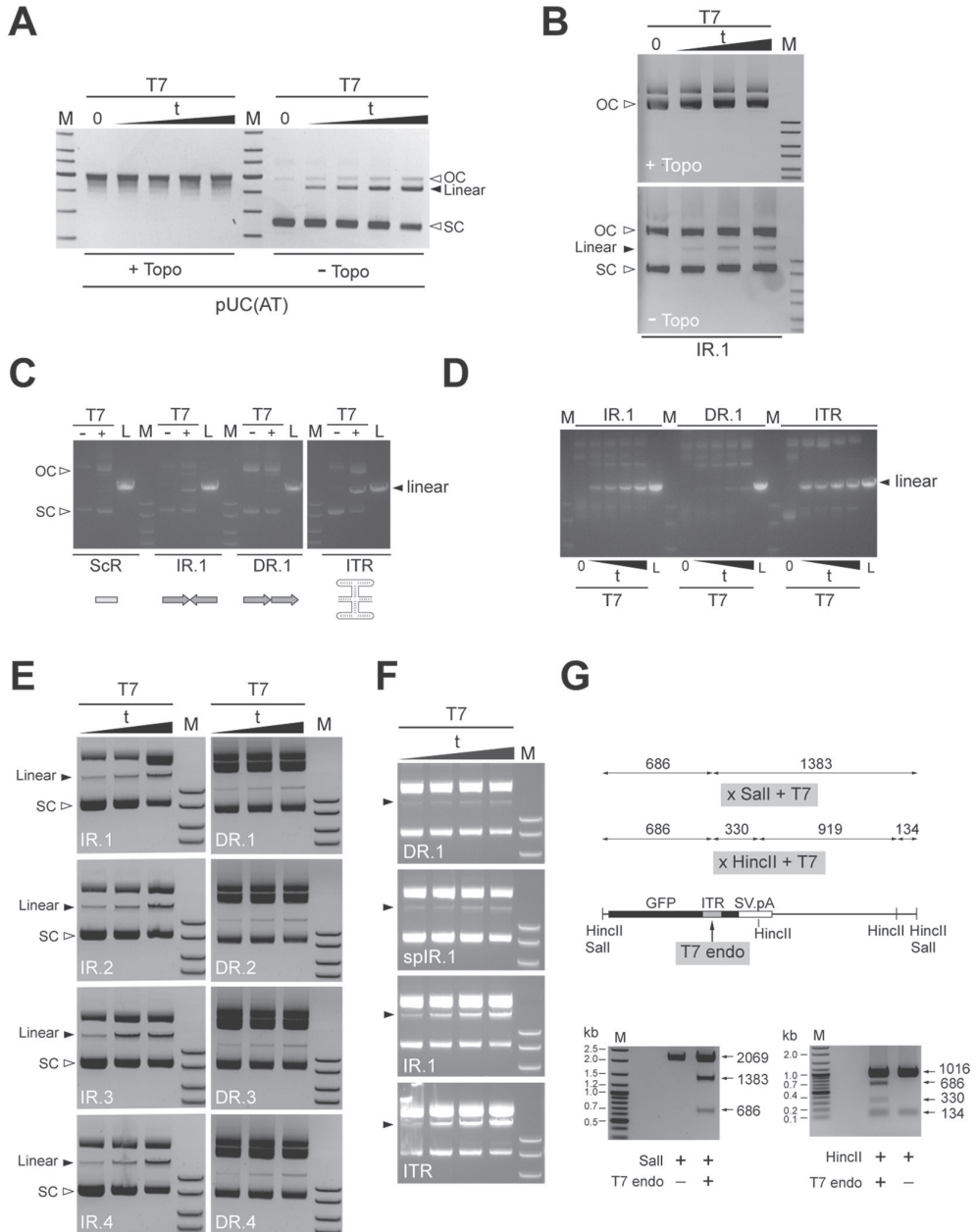
In another set of experiments, we investigated the T7 endonuclease I-dependent accumulation of linear acceptor DNA species as a function of time. To this end, constructs acceptor$^{IR.1}$, acceptor$^{DR.1}$ and acceptor^{ITR} were incubated with the resolvase for 10, 20, 30 or 60

min or left untreated. Analysis of the resulting DNA products by agarose gel electrophoresis showed that for each period of T7 endonuclease I treatment, the inverted repeat-containing plasmids acceptor[IR.1] and acceptor[ITR] yielded higher amounts of linear DNA than the direct repeat-containing construct acceptor[DR.1] (Figure 2D). In fact, exposure of acceptor[DR.1] for 60 min to T7 endonuclease I generated less linear DNA molecules than a 10-min incubation of acceptor[ITR] and acceptor[IR.1] with the same enzyme (Figure 2D). However, linear acceptor[ITR] molecules accumulated with faster kinetics than linear acceptor[IR.1] DNA (Figure 2D). These data, in agreement with the calculated folding free energy values (Figure 1C), indicate that T7 endonuclease I-susceptible secondary DNA structures (e.g. DNA cruciforms) can be formed after insertion of the test DNA sequences as inverted repeats in acceptor plasmid substrates (Figure 1C). These experimental results further suggest that the AAV ITR has a higher propensity to acquire a T7 endonuclease I-sensitive DNA conformation than IR.1. Similar experiments were also carried out with not only the acceptor substrates harboring IR.1, DR.1 and ITR but also with those containing each of the other 6 non-spaced repeat sequences (i.e. DR.2, DR.3, DR.4, IR.2, IR.3 and IR.4) as well as that with the spaced inverted repeat sequence spIR.1. Results from these experiments established that for each of the IR/DR acceptor pairs those substrates containing the non-spaced inverted repeat led, at very time point analyzed, to a higher accumulation of T7 endonuclease I-resolved linear forms than that resulting from those harboring the non-spaced direct repeat (Figure 2E, compare each left panel with the corresponding right panel). Of note, acceptor[spIR.1] with its spaced inverted repeat was clearly less prone to T7 endonuclease I digestion than its acceptor[IR.1]counterpart (Figure 2F, compare second with third panel from the top). Once again, acceptor[ITR] with its multi-palindromic AAV ITR sequence displayed a somewhat higher susceptibility to the T7 endonuclease I when compared to that of acceptor[IR.1] (Figure 2F, compare the two lowest panels). Finally, to identify the position corresponding to the major T7 endonuclease I cleavage site in acceptor DNA backbones with secondary structure-forming test sequences, we incubated an ITR-containing construct (Figure 2G, upper panel) exclusively with SalI or with SalI and T7 endonuclease I or exposed it to HincII or to HincII and T7 endonuclease I. Agarose gel electrophoresis of the resulting DNA fragments revealed digestion patterns fully consistent with T7 endonuclease I-dependent cleavage of the acceptor templates at the location of the inverted repeat (Figure 2G, lower panels).

Inverted repeats stimulate DNA exchange through homologous recombination in mammalian cells

Next, we deployed the aforementioned episomal HR assay system (Figure 1) to ask whether DNA templates containing direct or inverted DNA repeats can serve as substrates for homology-directed gene repair in mammalian cells. In these experiments, HeLa cells were either mock-transfected or transfected with different plasmid combinations. At

four days post-transfection, HR-dependent *GFP* repair was measured by flow cytometry. Transfection of the donor construct GFP[ΔATG], acceptor[DR.1] or acceptor[IR.1] alone did not give rise to measurable GFP-specific signals showing that the mutations introduced in these plasmids did functionally disrupt the *GFP* ORF (Figure 3A). Co-transfection of both GFP[ΔATG] and acceptor[ScR] resulted in a low percentage of GFP-positive cells (i.e. 0.3 ± 0.2 % [n=8]; Figure 3A, donor + ScR), establishing the background of the assay in the presence of both donor and acceptor templates. To validate the extrachromosomal recombination system, we relied on I-SceI-mediated site-specific DSB formation, which is a well-established method to induce HR in a controlled and predictable manner (see e.g. [19]). For this purpose, HeLa cells were co-transfected with GFP[ΔATG], acceptor[ScR] and the I-SceI-encoding expression plasmid pCAG.I-SceI. The inclusion of pCAG.I-SceI resulted in a large increase in the frequency of GFP-positive cells (i.e. 2.5 ± 0.4 % [n=11]; Figure 3A, donor + ScR + I-SceI). Interestingly, while co-transfection of GFP[ΔATG] and acceptor[DR.1] yielded background levels of GFP-positive cells (Figure 3A, compare donor + ScR with donor + DR.1), co-delivery of the donor construct together with acceptor[IR.1] gave rise to a significantly higher percentage of GFP-positive cells (i.e. 1.9 ± 0.5 % [n=11]; Figure 3A, compare donor + ScR and donor + DR.1 with donor + IR.1). These data imply that a tandem of test sequence 1 when arranged in an inverted repeat orientation serves as an effective target for homology-directed gene repair.

To further investigate the relationship between the relative orientation of repetitive DNA sequences and HR-mediated gene repair, HeLa cells were co-transfected with GFP[ΔATG] and with an acceptor plasmid containing a tandem of test sequences 2, 3 or 4 arranged in either a direct or in an inverted repeat configuration. Consistent with the previous results (Figure 3A), acceptor plasmids endowed with the various inverted repeats yielded significantly higher numbers of GFP-positive cells than their isogenic direct repeat-containing counterparts, which gave rise to frequencies of GFP-positive cells not significantly above background level (i.e. GFP[ΔATG] + acceptor[ScR]; Figure 3B). Thus, at least for the four different test DNA sequences investigated, which possess the same GC content, induction of HR is primarily dependent on the arrangement of the repetitive DNA unit as opposed to their specific nucleotide sequence.

Central spacing abolishes inverted DNA repeat-dependent homology-directed gene repair
To study the impact of repetitive DNA spacing in inverted repeat-induced HR, we generated acceptor[spIR.1]. This plasmid was made by inserting at the axis of symmetry of IR.1, a 42-bp sequence encompassing the ScR to effectively separate test sequence 1 from its reverse complement copy. In these experiments, HeLa cells were transfected with a mixture of the donor plasmid GFP[ΔATG], acceptor[ScR], acceptor[IR.1] or acceptor[spIR.1] and, where indicated, pCAG.I-SceI. As previously observed, co-transfection of GFP[ΔATG] and acceptor[IR.1] resulted in significant HR-dependent *GFP* expression (Figure 3C). However, substitution of acceptor[IR.1]

by acceptor$^{spIR.1}$, yielded *GFP* expression rescue activity levels that were not above those detected in cell cultures exposed to GFP^{ΔATG} and acceptor^{ScR} (Figure 3C). Possibly, physical separation of the repetitive DNA unit as in acceptor$^{spIR.1}$ inhibits the *in vivo* formation of cruciform-like structures as suggested by our *in vitro* assay results (Figure 2F) as well as those of others (22-24), which renders the transfected DNA templates no longer a target for the HR machinery. Addition of pCAG.I-SceI to the transfection mixtures consisting of GFP^{ΔATG} and either acceptor^{ScR} or acceptor$^{spIR.1}$ rescued high-level *GFP* expression (Figure 3C). The latter outcome shows that acceptor$^{spIR.1}$ does not contain an intrinsically defective *GFP* target template. Taken together, these data indicate that perfect palindromes or non-spaced inverted DNA repeats are preferred over spaced inverted DNA repeats as targets for homology-directed gene repair *in vivo* presumably due to their capacity to form secondary structures *in vivo* that can subsequently serve as direct targets for cellular structure-specific nucleases.

Figure 3. Effect of repetitive DNA sequences in a direct or inverted repeat configuration on homology-directed gene repair in HeLa cells. (**A**) HeLa cells were transfected with acceptor$^{DR.1}$ or acceptor$^{IR.1}$ alone (DR.1 and IR.1, respectively) or with either of these two acceptor plasmids in combination with the donor construct GFP^{ΔATG} (donor + DR.1 and donor + IR.1, respectively). Mock-transfected HeLa cells (mock) and HeLa cells transfected with GFP^{ΔATG} alone (donor) or together with acceptor^{ScR} (donor + ScR) served as negative controls. The positive control for the rescue of *GFP* expression by HR was provided by co-transfecting HeLa cells with acceptor^{ScR}, GFP^{ΔATG} and the I-SceI-encoding plasmid pCAG.I-SceI (donor + ScR + I-SceI). Quantification of the number of GFP-positive cells was carried out by flow cytometry at 4 days post-transfection. A minimum of 5 and a maximum of 11 independent experiments were performed with 10,000 events corresponding to viable cells being measured per sample. (**B**) HeLa cells were co-transfected with GFP^{ΔATG} plus either acceptor$^{IR.2}$, acceptor$^{DR.2}$, acceptor$^{IR.3}$, acceptor$^{DR.3}$, acceptor$^{IR.4}$ or acceptor$^{DR.4}$. To facilitate comparison, data sets corresponding to HeLa cells co-transfected with GFP^{ΔATG} and acceptor^{ScR} or with GFP^{ΔATG}, acceptor^{ScR} and pCAG.I-SceI as well as those corresponding to HeLa cells co-transfected with GFP^{ΔATG} and either acceptor$^{IR.1}$ or acceptor$^{DR.1}$ presented in Figure 3A are repeated in Figure 3B (open and grey bars, respectively). Quantification of *GFP* expression rescue was carried out by flow cytometry at 4 days post-transfection. Cumulative data from 4 different experiments (solid bars) are expressed as mean ± standard deviation. *p=0.002; **p=0.009. (**C**) Flow cytometric analysis of HeLa cells that, in addition to being exposed to the donor plasmid GFP^{ΔATG} also received acceptor^{ScR}, acceptor$^{IR.1}$ or acceptor$^{spIR.1}$. In the latter construct, the inverted repeat of test sequence 1 is interrupted at its axis of symmetry by an I-SceI recognition site (see diagram below the graph). HeLa cells co-transfected with GFP^{ΔATG}, the I-SceI encoding plasmid pCAG.I-SceI and either acceptor^{ScR} or acceptor$^{spIR.1}$ served as positive controls for HR-mediated *GFP* repair. Data corresponding to a minimum of 3 different experiments are shown as mean ± standard deviation. *p=0.0002; **p=0.005. ▶
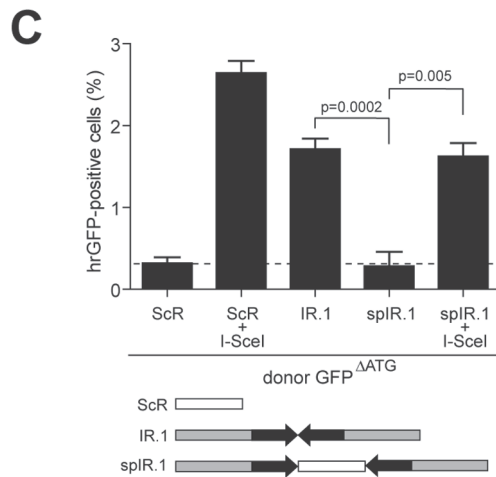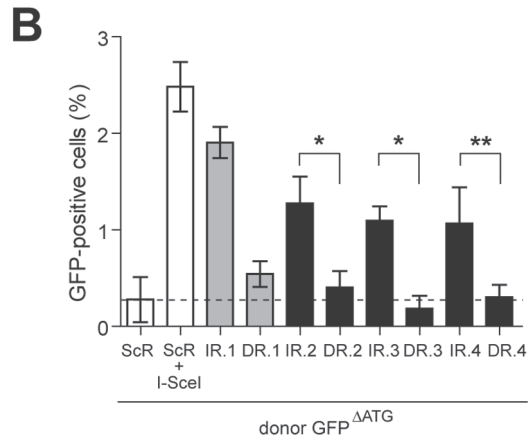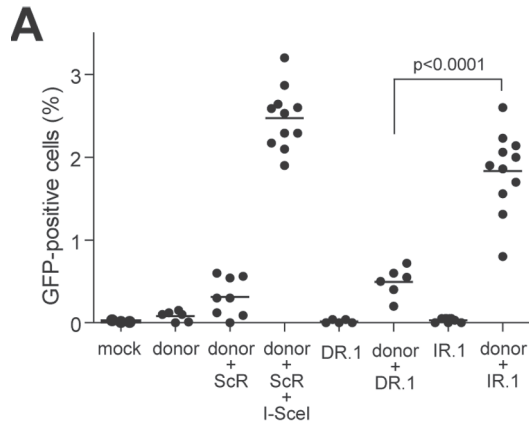
Experimental evidence for *in vivo* nuclease-mediated processing of non-spaced inverted DNA repeats

In the search of evidence for nuclease-mediated processing of palindromic test sequences *in vivo*, we set-up the assay system illustrated in Figure 4A. In this assay, HeLa cells are transfected with acceptor[IR.1], acceptor[DR.1] or with acceptor[ScR] mixed with pCAG.I-SceI or with pCAG.I-SceI(Δ112-246). Expression plasmid pCAG.I-SceI(Δ112-246) encodes a non-functional I-SceI protein. Cells co-transfected with acceptor[ScR] and pCAG.I-SceI constitute a positive control for *in vivo* site-specific DSB formation at acceptor templates. A key feature of this assay is the fact that a discriminating marker in the form of a BcuI recognition site lies at the axis of symmetry of the test non-spaced inverted repeat sequence (Figure 1B and Figure 4A, upper panel). Generation of cruciforms at this sequence followed by its recognition and processing by cellular structure-specific nuclease(s) should result in DNA breaks. The resulting DNA can subsequently serve as a substrate for error-prone DNA repair processes in the cell, such as non-homologous end-joining (NHEJ), eventually leading to the emergence of a population of BcuI-resistant acceptor molecules (Figure 4A). Likewise, processing of I-SceI-mediated DSBs by a cellular error-prone DNA repair pathway should yield templates that are knocked-out in the I-SceI cognate target site. Thus, DNA processing/repair at specific test sequences should lead to a mixture of *hrGFP* templates that can be PCR-amplified and discriminated on the basis of sequence-specific enzymatic digestions combined with Southern blot and nucleotide sequence analysis. Southern blotting of amplicons made with the aid of primer set 1/2 (Figure 4A), showed the presence of I-SceI-undigested templates in extrachromosomal DNA isolated from HeLa cells co-transfected with acceptor[ScR] and pCAG.I-SceI (+I-SceI; Figure 4B) but not in those co-transfected with acceptor[ScR] and pCAG.I-SceI(Δ112-246) (-I-SceI; Figure 4B). Importantly, the same analysis applied to episomal DNA extracted from HeLa cells transfected with acceptor[IR.1] or with acceptor[DR.1] revealed the presence of BcuI-undigested templates in cells exposed to the former construct (Figure 4B, lower-left panel). The fact that PCR products amplified from acceptor[ScR] or from acceptor[IR.1] and treated with XmaJI did not yield discernable undigested material suggests that the majority of the DNA sequence modifications took place in the vicinity of the respective I-SceI site or of the IR.1 axis of symmetry. Next, amplicons isolated from the BcuI-undigested fraction were cloned into pCR4-TOPO and subjected to restriction fragment length analysis using BcuI. This analysis confirmed the disruption of the BcuI restriction site in several of the analyzed clones (Figure 4C, lanes 1, 2, 4, 5 and 6). Moreover, the apparently different molecular weights of the restriction fragments migrating slower than 1.2 kb-sized linear DNA suggest that the BcuI-refractory clones do harbor sequences representing the end-product of independent DNA processing/repair events. To confirm this, and to identify the breakpoints in acceptor[IR.1] templates at the nucleotide level, we carried out DNA sequence analysis of clones 5 and 6 (Supplementary Figure S1). As a control, we

also sequenced a pCR4-TOPO-based clone corresponding to PCR-amplified DNA from HeLa cells co-transfected with acceptor[ScR] and pCAG.I-SceI (Supplementary Figure S1).
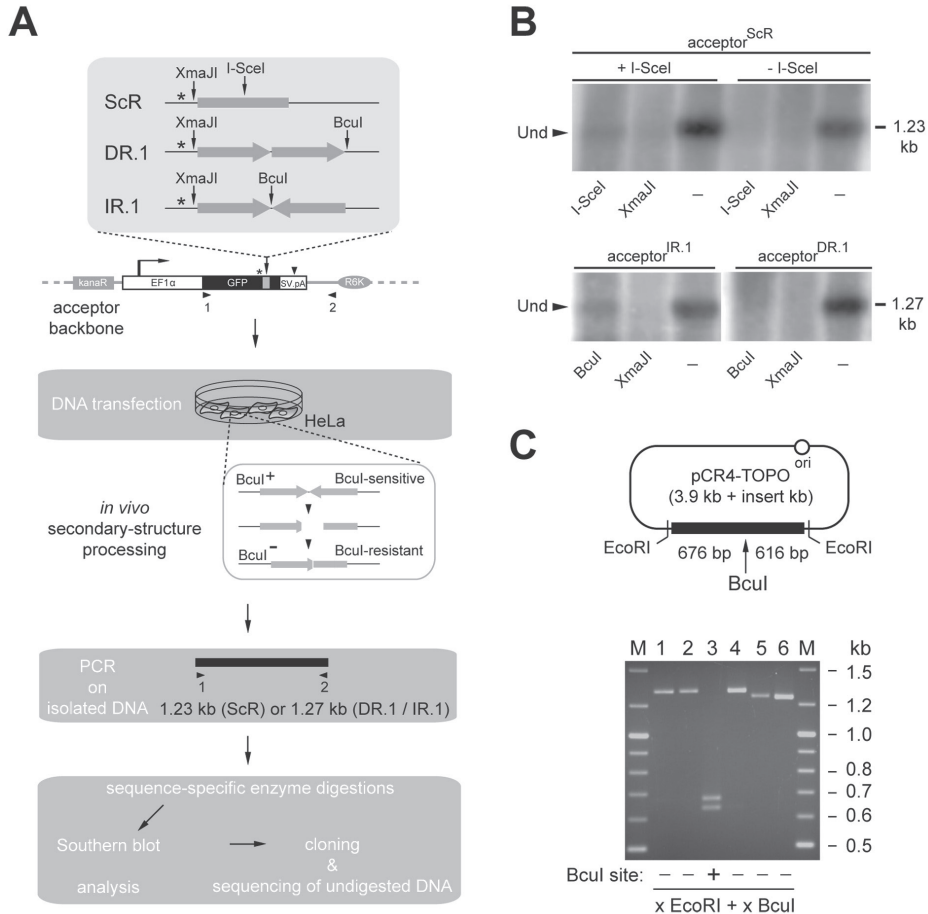
Figure 4. *In vivo* processing and repair of non-spaced inverted DNA repeats. (**A**) Diagrammatic representation of the experimental set-up deployed to study the processing and repair of non-spaced inverted DNA repeats in cells (see text for details and legend of Figure 1A for an explanation of the symbols). (**B**) Southern blotting analysis using an *hrGFP* ORF- and SV40 pA-specific probe of PCR products derived from extrachromosomal DNA isolated from HeLa cells co-transfected with acceptor[ScR] and pCAG-I.SceI (+I-SceI), acceptor[ScR] and pCAG-I.SceI(pCAG.I-SceI(Δ112-246) (-I-SceI), acceptor[IR.1] or acceptor[DR.1]. Prior to electrophoresis, the DNA samples were treated with DpnI alone (-) or with DpnI together with I-SceI, XmaJI or BcuI. Sizes corresponding to undigested amplicons (solid arrowhead) expected for the primer pair 1/2 are indicated on the right of the autoradiograms. (**C**) Upper panel, Diagram of a pCR4-TOPO molecular clone containing a PCR product amplified from extrachromosomal DNA isolated from HeLa cells transfected with acceptor[IR.1] and treated with BcuI. Vertical arrow points to the original position of the BcuI recognition site. Lower panel, Agarose gel electrophoresis of BcuI-treated pCR4-TOPO clones harboring PCR products amplified from episomal DNA extracted from HeLa cells transfected with acceptor[IR.1] and exposed to BcuI digestion (lanes 1 through 6). Lane M, GeneRuler DNA Ladder Mix molecular weight marker.

Inverted DNA repeat-dependent homology-directed gene repair occurs in a variety of mammalian cell types

Subsequently, we exposed cultures of HEK 293T cells, human fetal retinoblasts (911 and PER.tTA.Cre76 cells) and African green monkey kidney fibroblasts (COS-7 cells) to GFP$^{\Delta ATG}$ and either acceptor$^{DR.1}$ or acceptor$^{IR.1}$. Again, negative and positive controls were provided by co-transfecting cultures of each of these cell types with a mixture of the donor plasmid GFP$^{\Delta ATG}$ and acceptor^{ScR} alone or together with pCAG.I-SceI, respectively. Data depicted in Figure 5A, show distinct levels of HR-dependent *GFP* repair in the various cell types tested. This might be the result of different transfection efficiencies and/or of intrinsic cell type-specific differences in the ability to recognize and process, via HR, DNA secondary structures. Importantly, however, like in HeLa cells, appreciable HR-mediated *GFP* repair was only observed after co-transfection of GFP$^{\Delta ATG}$ and acceptor$^{IR.1}$ (Figure 5A). This IR.1-mediated HR stimulatory effect was independent of the amount of p53 present in the disparate cell types tested (Supplementary Figure S2). Some representative flow cytometry dot plots and direct fluorescence microscopy micrographs corresponding to these experiments are depicted in Figure 5B and Figure 5C, respectively. Collectively, these experiments suggest that inverted DNA repeat-induced HR is a mammalian cell type-independent phenomenon.

Figure 5. Effect of repetitive DNA sequences in a direct or inverted repeat configuration on homology-directed gene repair in mammalian cells. (**A**) Flow cytometric analysis of human embryonic kidney 293T cells (HEK), human fetal retinoblasts (911 and PER.tTA.Cre76 [PER]) and African green monkey kidney fibroblasts (COS-7) co-transfected with GFP$^{\Delta ATG}$ and either acceptor$^{DR.1}$ or acceptor$^{IR.1}$. Negative and positive controls for HR-mediated *GFP* repair in each of the tested cell types were provided by cells containing GFP$^{\Delta ATG}$ and acceptor^{ScR} (-) or these two plasmids as well as pCAG.I-SceI (+), respectively. (**B**) Dot plot representation of *GFP* expression in human embryonic kidney 293T cells (HEK) transfected with GFP$^{\Delta ATG}$ alone (donor) or with a mixture of GFP$^{\Delta ATG}$ and either acceptor$^{DR.1}$ or acceptor$^{IR.1}$. Cultures co-transfected with GFP$^{\Delta ATG}$, acceptor^{ScR} and the I-SceI expression plasmid pCAG.I-SceI served as positive control. Flow cytometry was carried out 3 days post-transfection with 10,000 viable cells being analyzed per sample. (**C**) Live-cell imaging by phase-contrast and fluorescence microscopy of monolayers of African green monkey kidney fibroblasts (COS-7) co-transfected with GFP$^{\Delta ATG}$ and either acceptor$^{IR.1}$ or acceptor$^{DR.1}$. Parallel cultures exposed to GFP$^{\Delta ATG}$, acceptor^{ScR} and pCAG.I-SceI served as positive control for HR-dependent *GFP* reconstitution. Microscopic analysis was performed 3 days post-transfection. Original magnification: 40×. ▶

A

B

HEK

C

COS-7
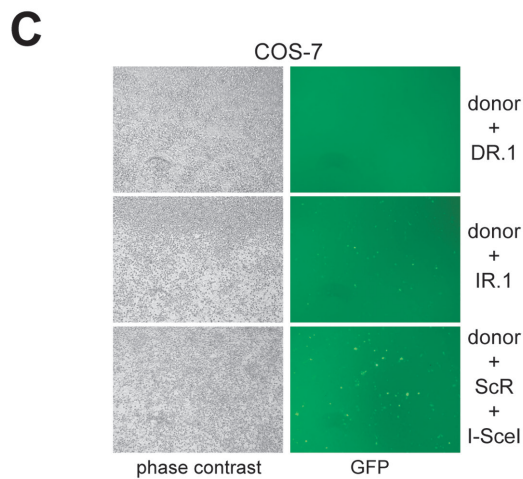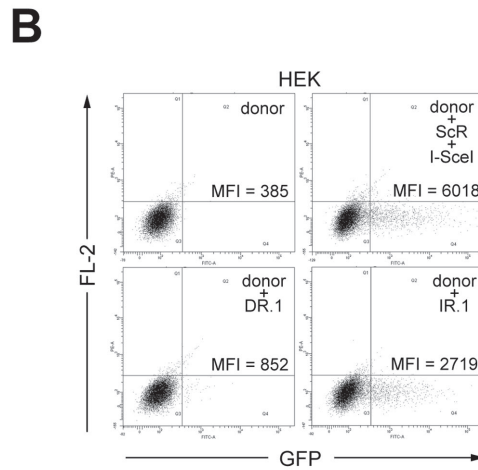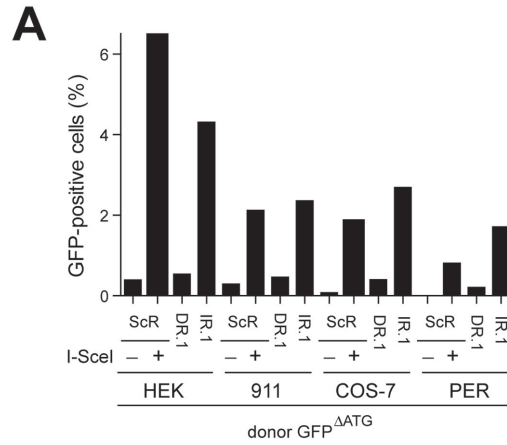
Composite inverted DNA repeats are equally effective at stimulating homology-directed gene repair as simple palindromes

To investigate the capacity of a structurally more "complex" or composite inverted DNA repeat to stimulate HR, we deployed acceptor[ITR]. Contrary to HeLa cell cultures that were transfected with donor GFP[ΔATG] or acceptor[ITR] alone, those exposed simultaneously to GFP[ΔATG] and acceptor[ITR] readily revealed the presence of GFP-positive cells (Figure 6A). The percentages of these cells were similar to those observed in HeLa cell cultures upon co-transfection of GFP[ΔATG] and acceptor[IR.1] (Figure 3). To rule out the possibility that DNA sequences prone to secondary structure formation, such as the AAV ITR, alter plasmid DNA transfection efficiency, we transfected HeLa cells with pDsRed or with pDsRed[2X.ITR]. The former construct contains an expression unit based on the human *elongation factor 1α* promoter, the *DsRed.T4* reporter and the rabbit *β-globin* polyadenylation signal, whereas the latter has this transcription unit flanked by AAV ITRs. Flow cytometric analysis at 72 hours post-transfection showed that cultures exposed to pDsRed[2X.ITR] had a frequency of reporter-positive cells as well as an amount of DsRed.T4 protein very similar to those measured in cultures transfected with pDsRed. We conclude that the ITR sequences do not significantly affect the transfection efficiency of plasmid DNA (Supplementary Figure S3).

To confirm, at the molecular level, the accurate repair of *GFP* ORFs following inverted repeat-mediated HR, we performed PCR analysis on extrachromosomal DNA isolated from HeLa cells transfected with GFP[ΔATG] alone or together with acceptor[ITR]. Extrachromosomal DNA extracted from HeLa cells exposed to acceptor[ITR] in combination with the non-homologous donor construct RFP[ΔATG] served as an extra negative control. The PCR assay shown in Figure 6B is based on primer pairs 1/2 and 1/3 to amplify PCR products that are diagnostic for the generation of *GFP* templates corrected by two-sided and one-sided HR, respectively. Although primers 1 and 2 also bind to the acceptor plasmid, their use did not yield any PCR products (Figure 6C) most likely due to the inability of the thermostable polymerase to the read through the AAV ITR. As shown in Figure 6C (upper and middle panels), PCR fragments corresponding to reconstituted *GFP* ORFs were exclusively detected in the DNA sample from the cells that were co-transfected with donor GFP[ΔATG] and acceptor[ITR]. Moreover, the results demonstrate that gene repair was brought about by two-sided as well as by one-sided HR (Figure 6C, lane 1 of upper panel and lane 1 of middle panel, respectively). Of note, amplification reactions carried out on an *in vitro* mixture of GFP[ΔATG] and acceptor[ITR] plasmids did not yield any product, showing that the detection of specific amplicons in the DNA sample derived from HeLa cells co-transfected with GFP[ΔATG] and acceptor[ITR] was not an artifact but the result of genetic information exchange *in vivo* (not shown). Internal control PCR amplifications using primers 4 and 5 showed the presence of the homologous and the non-homologous donor templates GFP[ΔATG] and RFP[ΔATG], respectively, confirming the integrity of the extrachromosomal DNA following the isolation procedure (Figure 6C, lower panel).

Next, we performed an independent transfection experiment in HeLa cells followed by the same DNA isolation procedure and PCR assay. In this new experiment, however, an extra control was included. This consisted in using extrachromosomal DNA from cells transfected with acceptor[ITR] mixed, prior to PCR, with extrachromosomal DNA from cells transfected with donor GFP[ΔATG]. Data depicted in Supplementary Figure S4 shows, once again, the presence of the specific 1.2-kb amplicon exclusively in the sample corresponding to cells co-transfected with donor[ΔATG] and acceptor[ITR].

The PCR products obtained with the aid of primer pair 1/2 were inserted into a plasmid vector after which, nucleotide sequence analysis of twenty randomly selected DNA clones was carried out. From this analysis resulted that nineteen of these clones contained *GFP* ORFs without any mutations linking them to error-free HR events (Figure 6D, 5 uppermost nucleotide sequences). The remaining clone had the *GFP* ORF disrupted at the initially engineered premature stop codon and AAV ITR-derived, heterologous, DNA (Figure 6D; lowest nucleotide sequence) suggesting that it was the product of inverted repeat microhomology-directed recombination (25) or of error-prone non-homologous end-joining (NHEJ), possibly following center-break palindrome revision or cruciform resolution (2,3,26,27). Additionally, we studied the kinetics of homology-directed gene repair involving acceptor[ITR] and acceptor[IR.1] and directly compared it to that of conventional DSB-induced HR. To this end, HeLa cells were transfected with donor GFP[ΔATG] alone or together with acceptor[IR.1], acceptor[ITR] or a mixture of acceptor[ScR] and pCAG.I-SceI. Results shown in Figure 6E, reveal a time-dependent increase in the number of GFP-positive cells (Figure 6E, upper graph) as well as in the amount of reporter protein per GFP-positive cell (Figure 6E, lower graph) in all cultures co-transfected with acceptor plasmids and the donor GFP[ΔATG] construct. Interestingly, the time-dependent increase in the frequency and fluorescence intensity of GFP-positive cells was faster in cultures exposed to acceptor[ScR] and pCAG.I-SceI than in those incubated with acceptor[ITR] or with acceptor[IR.1]. Moreover, no significant differences in both of the GFP-specific parameters were found at all time points tested in cell cultures exposed to GFP[ΔATG] and either acceptor[ITR] or acceptor[IR.1] (Figure 6). We postulate that the lower HR-inducing activity of the inverted DNA repeat sequences as compared to DSBs may relate to their transient nature. Perhaps, secondary structures formed *in vivo* by certain inverted repeats and palindromes constitute "facultative" or "intermittent" DNA lesions leading to a sporadic engagement of the HR machinery. Another contributing factor may be the larger number of biochemical reactions necessary to process cruciform-like structures by HR than that that is necessary to repair DSBs.

Figure 6. Comparison of the ability of simple and composite inverted DNA repeats to trigger homology-directed gene repair in mammalian cells. (**A**) HeLa cells were transfected with GFP$^{\Delta ATG}$ (donor) or with acceptor^{ITR} (ITR) or were co-transfected with GFP$^{\Delta ATG}$ plus acceptor^{ITR} (donor + ITR). Analysis of *GFP* expression by flow cytometry was carried out 4 days post-transfection on 10,000 viable cells per sample. Horizontal lines representing means were derived form 9 independent experiments. (**B**) Schematic representation of the PCR-based assay deployed to detect *GFP* ORFs repaired via homology-directed gene targeting. Primers 1 and 2 were designed to amplify templates resulting from two-sided HR whilst oligodeoxyribonucleotides 1 and 3 were used to specifically detect products of one-sided HR. Primer 1 recognizes the first 24 nts of the *GFP* ORF. Oligodeoxyribonucleotides 2 and 3 target sequences exclusively present in the acceptor and donor plasmid backbones, respectively. The PCR products of primers 4 and 5, which bind to the rabbit *β-globin* pA (βG.pA) and SV40 pA (SV.pA), respectively, served as internal control for extrachromosomal DNA quality and quantity. Amplicon size (in bp) expected for each primer pair is indicated. For an explanation of the other symbols see the legend of Figure 1A. (**C**) PCR analysis using primer pairs 1/2, 1/3 and 4/5 (upper, middle and lower panels, respectively) of extrachromosomal DNA isolated from HeLa cells co-transfected with acceptor^{ITR} and the homologous donor plasmid GFP$^{\Delta ATG}$ (lane 1) or with acceptor^{ITR} and the non-homologous donor plasmid RFP$^{\Delta ATG}$ (lane 2) or from HeLa cells transfected with GFP$^{\Delta ATG}$ alone (lane 3). Lane M, GeneRuler DNA Ladder Mix molecular weight marker. (**D**) Nucleotide sequence data of individual clones corresponding to PCR products obtained with primers 1 and 2. Clones #5, #9, #14, #16 and #18 represent products of HR containing a repaired *GFP* ORF. The encircled ORF-correcting cytosine is derived from the donor plasmid. Clone #4 corresponds to a rearranged acceptor template featuring 38 bp of the originally introduced AAV ITR (purple line above graph) and retaining the engineered stop codon. The G marked with the asterisk is derived from the original acceptor template. (**E**) Flow cytometric analysis at the indicated time points of HeLa cells transfected with the donor construct GFP$^{\Delta ATG}$ alone or together with either acceptor$^{IR.1}$, acceptor^{ITR} or a mixture of acceptor^{ScR} and the I-SceI-encoding plasmid pCAG.I-SceI. Both the frequencies of GFP-positive cells (upper graph) as well as their average mean fluorescence intensities (MFI; lower graph) are presented. ▶
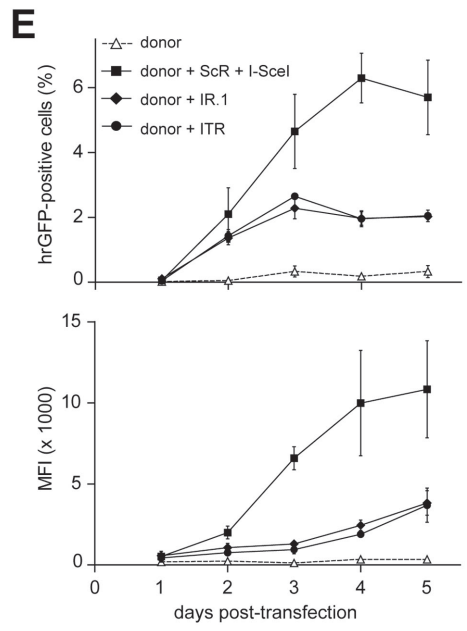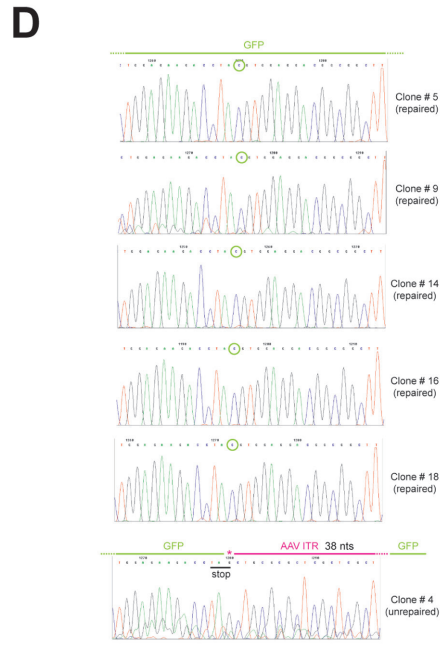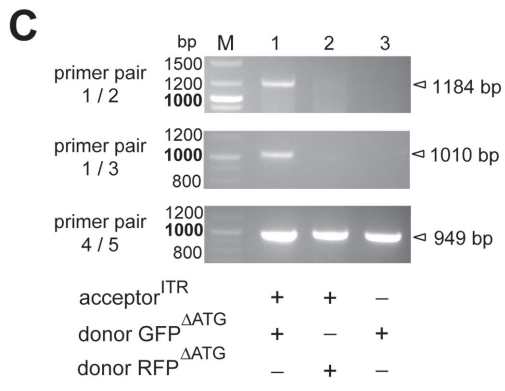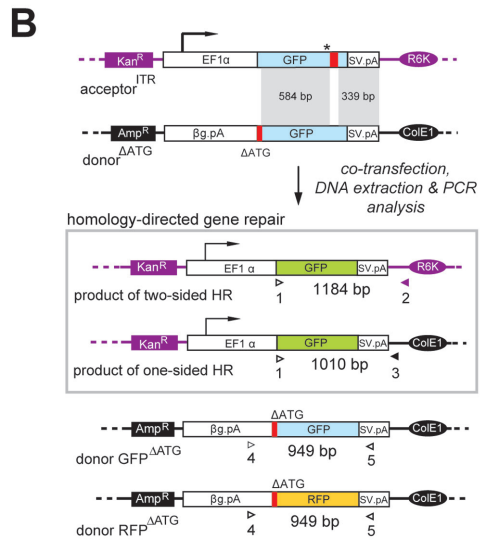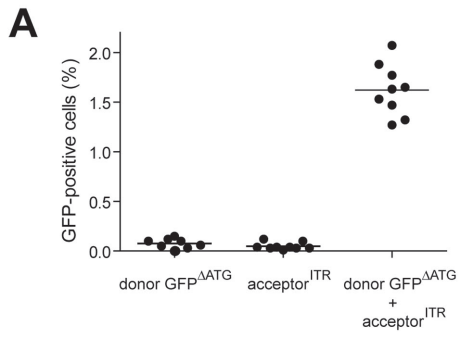
Replication of DNA repeat-containing acceptor molecules does not significantly affect homology-directed gene repair

SV40 is a mammalian double-stranded DNA virus with a circular genome whose replication has been extensively studied as a model for chromosomal nuclear DNA replication in higher eukaryotes (28). The only viral *cis*-acting element and *trans*-acting factor required for SV40-dependent DNA replication are the ori and the large T antigen protein, respectively. Thus, to investigate the impact of target template replication on DNA repeat-induced HR in mammalian cells, the SV40 ori was introduced at equivalent positions in acceptor[ScR], acceptor[DR.1], acceptor[IR.1] and acceptor[spIR.1] to generate the constructs acceptor[ScR.ORI], acceptor[DR.1.ORI], acceptor[IR.1.ORI] and acceptor[spIR.1.ORI], respectively. Next, each of these plasmids was individually transfected into the SV40 large T-expressing COS-7 cells together with the homologous donor construct GFP[ΔATG] or the non-homologous donor plasmid RFP[ΔATG]. Extrachromosomal DNA isolated from these cells was treated with DpnI to selectively digest the input prokaryotic DNA and with XbaI to linearize *de novo* synthesized acceptor DNA molecules and HR products. Southern blot analysis of the digestion products using a *GFP*-specific probe revealed SV40 ori-dependent accumulation of *de novo* generated DNA molecules, demonstrating the replication proficiency of the SV40 ori-containing acceptor plasmids in COS-7 cells (Figure 7A). Interestingly, acceptor DNA replication did not lead to a significant increase in homology-directed gene repair levels in any of the experimental set-ups tested (Figure 7B). Taken together, these data indicate that under the prevailing experimental conditions, replication of acceptor DNA molecules carrying a palindrome, a direct repeat or a spaced inverted repeat does not significantly enhance homology-directed gene repair.
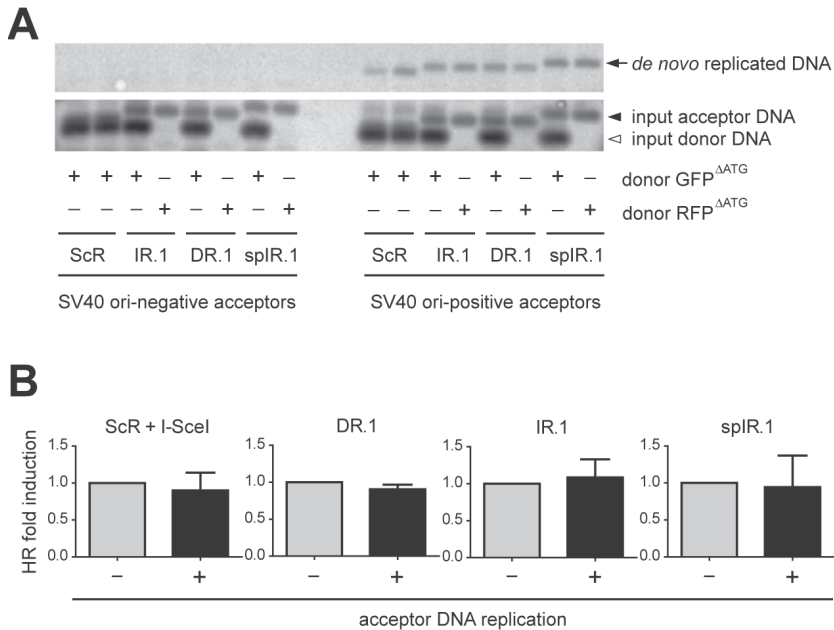
Figure 7. Testing the impact of target DNA synthesis on DNA repeat-mediated homology-directed gene repair. (**A**) SV40 ori-dependent DNA replication of acceptor constructs. Acceptor plasmids containing the test sequences ScR, IR.1, DR.1 or spIR.1 and with or without SV40 ori were transfected into COS-7 cells together with the homologous donor construct GFP$^{\Delta ATG}$ or the non-homologous donor plasmid RFP$^{\Delta ATG}$. At 3 days post-transfection, extrachromosomal DNA was extracted and treated with XbaI and the prokaryotic DNA methylation pattern-sensitive restriction enzyme DpnI. After agarose gel electrophoresis, the resolved DNA was subjected to Southern blot analysis using a *GFP*-specific probe. DpnI-resistant, *de novo* replicated DNA, was detected only in samples of cells transfected with SV40 ori-positive acceptor plasmids (right-hand side upper panel). (**B**) Relative homology-directed gene repair frequencies in COS-7 cells transfected with GFP$^{\Delta ATG}$, the indicated acceptor plasmids with (+) or without (-) SV40 ori and in one case also pCAG.I-SceI.

DISCUSSION

Repetitive DNA sequences include not only single, direct and inverted repeats (2-4), like the ones investigated in this study, but also high-copy-number repetitive DNA tracts such as those corresponding to 1-4 bp microsatellites and 6-64 bp minisatellites (1). Microsatellites include the expandable trinucleotide repeats associated with neurodegenerative and neuromuscular disorders such as Huntington's and oculopharyngeal muscular dystrophy. Despite their diversity, diverse lines of evidence point to the acquisition of non-B conformations by DNA at these motifs (e.g. hairpins, cruciforms, G-quadruplex, Z-DNA and H-DNA) as a common culprit through which they exert their biological effects possibly in concert with DNA metabolic and other DNA-related processes (4). Indeed, an increasing

number of experiments mainly carried out in bacterial and yeast model systems indicates that long single DNA repeats (i.e. >150 bp; [3]) with the potential to form secondary structures (e.g. hairpins and cruciforms) can serve as targets for the shuffling and exchange of genetic information (2,3).

The knowledge about the biological activity of different types of DNA repeats in relation to gene repair pathways (especially single repeats in the size range <150 bp) and the putative role played in these processes by DNA replication is scant. This knowledge gap is particularly acute in cells of higher eukaryotes (2,3). In this study, we have devised an extrachromosomal functional read-out system based on pairs of complementary DNA templates carrying defective GFP-encoding sequences that can serve as substrates for intermolecular HR-dependent gene repair. This experimental system allowed us to investigate in a quantitative manner the effect of various types of single DNA repeats on the HR process in mammalian cells. Furthermore, by endowing acceptor DNA molecules with a eukaryotic origin of replication, we could probe in a strict manner the role of template DNA synthesis on repeat-induced homology-directed gene repair. We found that, contrary to direct and to spaced inverted repeats, both simple palindromes as well as composite inverted DNA repeats constitute targets for the HR pathway in mammalian cells. Induction of homology-directed gene repair was dependent on the arrangement and spacing of the repetitive DNA unit rather than on its nucleotide sequence. We also found that the presence of inverted DNA repeat sequences in target molecules rendered them susceptible to coordinated nicking by T7 endonuclease I, a *bona fide* four-way DNA branch resolving enzyme (29). These results are consistent with other *in vitro* data showing that lineform-to-cruciform transition in double-stranded DNA molecules relies on the presence of an inverted repeat and is negatively affected by intervening spacer sequences in a length-dependent manner (22-24). We thus demonstrate that non-spaced inverted DNA repeats *per se* can stimulate homology-directed gene repair in mammalian cells presumably due to their capacity to form secondary structures *in vivo* that can subsequently serve as direct targets for cellular structure-specific nucleases. These processes may eventually lead to the formation of DSBs that constitute canonical substrates for, amongst others, NHEJ- and HR-based allelic and non-allelic recombination. Indeed, Inagaki and co-workers have recently shown in 293 cells by using a two-plasmid system together with a PCR-based assay that large secondary structure-forming palindromic AT-rich repeats (PATRRs), often associated with translocations in the human germ line, stimulate intermolecular rearrangements via a pathway likely to involve NHEJ (27). The impact of template DNA replication on the PATRR-specific rearrangements was, however, not investigated.

Resolution of cruciform-like structures is thought to start with the introduction of single-strand breaks on opposite sites of the branch point followed by a ligation step resulting in the generation of hairpin-capped termini that can be further processed by nicking to

generate "open" ends. Candidate resolving and processing enzymatic activities are those of the first isolated *bona fide* mammalian Holliday junction resolvase Gen1 (30) and Mre11 (31), respectively. Other candidate resolvase is that corresponding to the SLX4 complex (32). Possibly outcomes of such ectopic recombination processes include chromosomal translocations and loss-of-heterozigosity. Related to this, *in silico* analysis of the human genome as well as experiments in yeast suggest that, during evolution, palindromes and inverted repeats with short spacers are counter selected compared to direct repeats and inverted repeats with long spacers (9). Indeed, without implying causality, more recent computer-aided phylogenetic sequence analyses revealed a correlation between DNA repeat pairs, NHEJ and non-allelic HR in the shaping of mammalian genome evolution (33).

Finally, we also showed that, at least under conditions that do not disrupt processivity of DNA synthesis, replication of molecules harboring the direct or the inverted DNA repeats did not significantly increase the frequencies of HR-dependent gene repair events when compared to those measured in the absence of acceptor DNA replication. This finding on single DNA repeats adds to recent results indicating that, at least in the case of the high-copy-number trinucleotide repeat associated with Friedreich's ataxia GAA×TTC, DNA rearrangements can ensue in the absence of replication (34). Other processes like DNA transcription and certain repair pathways such as the herein examined HR can provide for alternative mechanisms underlying DNA repeat-associated rearrangements as proposed elsewhere (1,35). Indeed, the fact that repeat-associated DNA instability can occur independently of a replication-based mechanism (e.g. replication stalling or slippage) can also be circumstantially inferred from the significant age-dependent expansion of secondary structure-forming trinucleotide repeats in post-mitotic neurons of patients afflicted by certain neurodegenerative disorders (1).

Current models of inverted repeat-driven secondary structure formation *in vivo* posit that palindromes or quasi-palindromes can, under torsional strain, transit from lineform to cruciform in double-stranded DNA via intrastrand annealing. On the other hand, spaced inverted repeats, albeit being also self-complementary, can only hybridize in the single-stranded form such as when a DNA replication fork advances through them and concomitantly gives rise to the Okazaki initiation zone (OIZ) in the lagging strand. Possibly, under these conditions, and depending on the length of the repeat/spacer sequences relative to that of the OIZ, lagging strand self-annealing becomes thermodynamically favorable resulting in the formation of hairpins that can stall DNA replication (2,36) Interestingly, we showed that replication of DNA molecules containing the spaced inverted repeat spIR.1 could not overcome their inability to promote homology-directed gene repair. Related to this, Voineagu and colleagues (36) have recently deployed a SV40 ori-based plasmid system and 2D agarose gel electrophoresis to demonstrate that *Alu*-derived 320-bp-long inverted repeats with no spacer or with a relatively short 12-bp spacer lead to replisome

stalling in COS-1 cells, whereas the same inverted repeat with a 52-bp spacer did not. The authors interpreted these results as a consequence of the OIZ size limit not allowing effective stem-loop hairpin formation by inverted DNA repeats with the larger spacer. Thus, on the basis of our results and those of Voineagu and co-workers and towards dissecting the inverted repeat parameters allowing hairpin assembly through the postulated lagging strand displacement-dependent mechanism, it will be interesting to evaluate different-sized repeat/spacer sequences and their relationship with replication fork stalling on one hand (36) and homology-directed gene repair on the other (this study). These experiments might help to define the rules underlying secondary structure formation, replisome stalling and the HR-inducing activity of inverted DNA repeats in mammalian cells. Finally, the functional genetic assay described herein might also be helpful in evaluating the effect of other types of DNA motifs/parameters on HR in higher eukaryotes and the contribution of cellular factors to this process.

REFERENCES

1. Pearson,C.E., Edamura,K.C. and Cleary,J.D. (2005) Repeat instability: Mechanisms of dynamic mutations. *Nat. Rev. Genet.*, **6**, 729-742.

2. Lewis,S.M. and Coté,A.G. (2006) Palindromes and genomic stress fractures: Bracing and repairing the damage. *DNA repair*, **5**, 1146-1160.

3. Lobachev,K.S., Rattray,A. and Narayanan,V. (2007) Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Front. Biosci.*, **12**, 4208-4220.

4. Zhao,J., Bacolla,A., Wang,G. and Vasquez,K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life. Sci.*, **67**, 43-62.

5. Kim,E.L., Peng,H., Esparza,F.M., Maltchenko,S.Z. and Stachowiak,M.K. (1998) Cruciform-extruding regulatory element controls cell-specific activity of the tyrosine hydroxylase gene promoter. *Nucleic Acids Res.*, **26**, 1793-1800.

6. Haber,J.E. and Debatisse,M. (2006) Gene amplification: Yeast takes a turn. *Cell*, **125**, 1237-1240.

7. Bikard,D., Loot,C., Baharoglu,Z. and Mazel,D. (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol. Mol. Biol. Rev.*, **74**, 570-588.

8. Eykelenboom,J.K., Blackwood,J.K., Okely,E. and Leach,D.R.F. (2008) SbcCD causes a double-strand break at a DNA palindrome in the Escherichia coli chromosome. *Mol. Cell*, **29**, 644-651.

9. Lobachev,K.S., Stenger,J.E., Kozyreva,O.G., Jurka,J., Gordenin,D.A. and Resnick,M.A. (2000) Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.*, **19**, 3822-3830.

10. Fallaux,F.J., Kranenburg,O., Cramer,S.J., Houweling,A., Van Ormondt,H., Hoeben,R.C. and van der Eb,A.J. (1996) Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors. *Hum. Gene Ther.*, **7**,215-222.

11. Gonçalves,M.A.F.V., van der Velde,I., Janssen,J.M., Maassen,B.T., Heemskerk,E.H., Opstelten,D.-J., Knaän-Shanzer,S., Valerio,D. and de Vries,A.A.F. (2002) Efficient generation and amplification of high-capacity adeno-associated virus/adenovirus hybrid vectors. *J. Virol.*, **76**, 10734-10744.

12. van Nierop,G.P., de Vries,A.A.F., Holkers,M., Vrijsen,K.R. and Gonçalves,M.A.F.V. (2009) Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Res.*, **37**, 5725-5736.

13. Yue,Y. and Dongsheng,D. (2002) Development of multiple cloning site cis-vectors for recombinant adeno-associated virus production. *Biotechniques*, **33**, 676-678.

14. Bevis,B.J. and Glick,B.S. (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat. Biotechnol.*, **20**, 83-87.

15. Richardson,C., Moynahan,M.E. and Jasin,M. (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev.*, **12**, 3831-3842.

16. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **313**, 3406-3415.

17. SantaLucia Jr.,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, **95**, 1460-1465.

18. Gonçalves, M.A.F.V., Pau,M.G., de Vries,A.A.F. and Valerio,D. (2001) 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.

19. Weinstock,D.M., Nakanishi,K., Helgadottir,H.R. and Jasin,M. (2006) Assaying double-strand break repair pathway choice in mammalian cells using a targeted endonuclease or the RAG recombinase. *Methods Enzymol.*, **409**, 524-540.

20. Ren,J., Qu,X., Chaires,J.B., Trempe,J.P., Dignam,S.S. and Dignam,J.D. (1999) Spectral and physical characterization of the inverted terminal repeat DNA structure from adenoassociated virus 2. *Nucleic Acids Res.*, **27**, 1985-1990.

21. Gonçalves, M.A.F.V. (2005) Adeno-associated virus: from defective virus to effective vector. *Virol. J.*, **2**, e43.

22. Vologodskii,A.V. and Frank-Kamenetskii,M.D. (1983) The relaxation time for a cruciform structure in superhelical DNA. *FEBS Lett.*, **160**, 173-176.

23. Sinden,R.R., Zheng,G.X., Brankamp,R.G. and Allen,K.N. (1991) On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. *Genetics*, **129**, 991-1005.

24. Kogo,H., Inagaki,H., Ohye,T., Kato,T., Emanuel,B.S. and Kurahashi,H. (2007) Cruciform extrusion propensity of human translocation-mediating palindromic AT-rich repeats. *Nucleic Acids Res.*, **35**, 1198-1208.

25. Kato,T., Inagaki,H., Kogo,H., Ohye,T., Yamada,K., Emanuel,B.S. and Kurahashi,H. (2008) Two different forms of palindrome resolution in the human genome: deletion or translocation. *Hum. Mol. Genet.*, **17**, 1184-1191.

26. Cunningham,L.A., Coté,A.G., Cam-Ozdemir,C. and Lewis,S.M. (2003) Rapid, stabilizing palindrome rearrangements in somatic cells by the center-break mechanism. *Mol. Cell. Biol.*, **23**, 8740-8750.

27. Inagaki,H., Ohye,T., Kogo,H., Kato,T., Bolor,H., Taniguchi,M., Shaikh,T.H., Emanuel,B.S. and Kurahashi,H. (2009) Chromosomal instability mediated by non-B DNA: Cruciform conformation and not DNA sequence is responsible for recurrent translocations in humans. *Genome Res.*, **19**, 191-198.

28. Fanning,E. and Zhao,K. (2009) SV40 DNA replication: from the A gene to a nanomachine. *Virology*, **384**, 352-359.

29. Declais,A.-C. and Lilley, D.M.J. (2008) New insight into the recognition of branched DNA structure by junction-resolving enzymes. *Curr. Opi. Struct. Biol.*, **18**, 86-95.

30. Ip,S.C.Y., Rass,U., Blanco,M.G., Flynn,H.R., Skehel,J.M. and West,S.C. (2008) Identification of Holliday junction resolvases from humans and yeast. *Nature*, **456**, 357-361.

31. Lobachev,K.S., Gordenin,D.A. and Resnick,M.A. (2002) The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell*, **108**, 183-193.

32. Svendsen,J.M. and Harper,J.W. (2010) GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. *Genes Dev.*, **24**, 521-536.

33. Zhao,H. and Bourque,G. (2009) Recovering genome rearrangements in the mammalian phylogeny. *Genome Res.*, **19**, 934-942.

34. Ditch,S., Sammarco,M.C., Banerjee,A. and Grabczyk,E. (2009) Progressive GAA×TTC repeat expansion in human cell lines. *PLoS Genet.*, **5**, e1000704.

35. Wang,G. and Vasquez,K.M. (2009) Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol. Carcinogen.*, **48**, 286-298.

36. Voineagu,I., Narayanan,V., Lobachev,K.S. and Mirkin,S.M. (2008) Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl. Acad. Sci. USA*, **105**, 9936-9941.

2

Differential integrity of *TALE* nuclease genes following adenoviral and lentiviral vector gene transfer into human cells

Maarten Holkers[1], Ignazio Maggio[1], Jin Liu[1], Josephine M. Janssen[1], Francesca Miselli[2], Claudio Mussolino[3], Alessandra Recchia[2], Toni Cathomen[3] and Manuel A.F.V. Gonçalves[1,]

[1]Department of Molecular Cell Biology, Leiden University Medical Center, Eithovenweg 20, 2333 ZC Leiden, The Netherlands. [2]Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 287, 41125 Modena, Italy. [3]Laboratory of Cell and Gene Therapy, Center for Chronic Immunodeficiency, University Medical Center Freiburg, Engesserstr. 4, 79108 Freiburg, Germany.

ABSTRACT

The array of genome editing strategies based on targeted double-stranded DNA break (DSB) formation have recently been enriched through the introduction of transcription activator-like type III effector nucleases (TALENs). To advance the testing of TALE-based approaches, it will be crucial to deliver these custom-designed proteins not only into transformed cell types but also into more relevant, chromosomally stable, primary cells. Viral vectors are among the most effective gene transfer vehicles. Here, we investigated the capacity of lentiviral and adenoviral vectors to package and deliver functional *TALEN* genes into various human cell types. To this end, we attempted at assembling particles of these two vector classes, each of which encoding a monomer of a TALEN pair targeted to a bipartite sequence within the *AAVS1* "safe harbor" locus. Vector DNA analyses revealed that adenoviral vectors transferred intact *TALEN* genes, whilst lentiviral vectors failed to do so, as shown by their heterogeneously sized proviruses in target cells. Importantly, adenoviral vector-mediated *TALEN* gene delivery resulted in site-specific DSB formation at the intended *AAVS1* target site at similarly high levels in both transformed and nontransformed cells. In conclusion, we demonstrate that adenoviral, but not lentiviral, vectors constitute a valuable *TALEN* gene delivery platform.

INTRODUCTION

The ability to design DNA-binding motifs that recognize pre-defined sequences of choice within complex eukaryotic genomes is opening a wide range of basic and applied research possibilities that involves the modulation of endogenous gene expression and chromosomal DNA editing. The latter aim has been mostly pursued by designing zinc-finger nucleases (ZFNs) comprising a sequence-specific array of synthetic zinc-finger motifs fused to the nuclease domain of the type IIS restriction endonuclease FokI (1,2). In addition, the recoding of natural homing endonucleases for heritable genome modification purposes has also been undergoing intense investigation (3,4). These artificial enzymes are tailored to generate double-stranded DNA breaks (DSBs) at pre-defined genomic sequences. As a consequence of their activity, targeted gene knock-out or chromosomal insertion of exogenous DNA can ensue following the activation and engagement of the error-prone non-homologous end-joining (NHEJ) or the error-free homologous recombination (HR) DNA repair pathways, respectively.

The recent discovery of the rules governing the recognition of specific DNA sequences by bacterial transcription activator-like type III effectors (TALEs) (5,6) has translated in a novel and straightforward methodology to construct proteins with customized DNA binding specificities. TALEs are proteins found in phytopathogenic bacteria of the genus *Xanthomonas*, whose main well-established function is to induce expression of specific host plant genes that enhance virulence. TALEs comprise an N-terminal translocation domain, a central region of tandem direct repeats that confers DNA binding specificity and a nuclear localization motif preceding a C-terminal transcriptional activation domain (7). The DNA binding domain of TALEs displays a unique architecture composed of a tandem array of typically 15.5 to 19.5 repeat units with *circa* 34 residues each (7). Often, the only distinguishing feature among the different repeats is a 2-amino acid hypervariable polymorphism at positions 12 and 13 dubbed "repeat-variable di-residue" (RVD). Crucially, it has been discovered that each individual RVD dictates the binding of the repeat in which it is embedded to a single nucleotide (5,6). Therefore, this one repeat to one nucleotide code directly establishes a simple rule governing TALE-DNA interactions. This insight combined with the modular nature of the DNA binding domain, is permitting researchers to harness TALE-derived repeats as custom-made scaffolds on which to assemble functional heterologous domains. Hitherto, designer TALEs tested *in vitro* and *in cellula* have included those with transcriptional activation or FokI nuclease domains (8,9). The latter, called TALE nucleases (TALENs) operate similarly to ZFNs in that a pair is assembled at a given DNA sequence consisting of two half-target sites separated by a spacer sequence (10). The directional binding of each TALEN monomer to its respective half-target site induces dimerization of the FokI portions resulting in site-specific DNA cleavage. Importantly, in contrast to zinc-finger modules (11), there are as of yet no

indications that binding of an individual TALE repeat to its cognate base pair is altered by neighboring sequences. This context-independent feature combined with the simple DNA binding code suggests that, when compared to ZFNs, TALENs can bind a wider range of DNA sequences and be programmed in an easier and more predictable manner. Moreover, the ability to construct functionally viable TALEN pairs with relatively large arrays of repeats has the potential to render them more specific than ZFNs.

Thus far, nucleic acids encoding TALENs have been introduced into cells by nucleic acid microinjection or by transfection methods based on chemical agents or electroporation (8,9,12-14). Unfortunately, these methods suffer from low throughput or are too cytotoxic or too inefficient in populations of non-transformed cells. Notably, the therapeutic application and the thorough assessment of the genome-wide TALEN specificity, for instance, will be dependent on highly efficient TALEN delivery systems into relevant karyotypically stable cells. In the realm of biomedical research this will include tissue-specific stem and progenitor cells. From the broad variety of methods to introduce foreign nucleic acids into cells, viral vectors have proven to be the most effective thus far. In particular, vectors derived from the double-stranded linear DNA human adenovirus serotype 5 and from the single-stranded positive-sense RNA lentivirus human immunodeficiency type 1 (HIV-1) have become broadly used (15,16). Due to the infection mechanisms evolved by their parental viruses (17,18), adenoviral and lentiviral vectors are particularly fit for the *in vitro* and *in vivo* transduction of dividing and non-dividing cells. In the current study, we investigated the suitability of these two major vector classes as delivery vehicles of *TALEN* sequences. We report that, in contrast to lentiviral vectors, adenoviral vectors can package and deliver functional TALEN-encoding expression units into transformed, immortalized and primary human cells.

MATERIALS AND METHODS

Cells

The 293T and HeLa cells (American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (DMEM; Invitrogen). The medium used for the culturing of the former and latter cell types was supplemented with 10% and 5% fetal bovine serum (FBS; Invitrogen), respectively. The adenoviral vector packaging cell lines PER.C6 (19) and PER.E2A (20) were maintained in DMEM containing 10% FBS, 10 mM MgCl2 (Sigma-Aldrich) without and with 250 μg/ml Geneticin (Invitrogen), respectively. These cells were kept in a humidified-air 10% CO2 atmosphere. The human myoblasts were obtained from a Duchenne muscular dystrophy patient and were immortalized through constitutive expression of *Bmi-1* and *TERT*. The generation of and culture conditions for these immortalized human myoblasts have been specified elsewhere (21). The human bone marrow-derived primary

fetal mesenchymal stem cells (hMSCs) were cultured as previously specified (22). The human myoblasts and hMSCs were kept in a humidified-air 5% CO_2 atmosphere.

Recombinant DNA

A TALEN pair (Supplementary Figure S1) targeting intron 1 of the *PPP1R12C* locus (*AAVS1* target site: 5'-TCTGTCCCCTCCACCCCAC-spacer-GACAGGATTGGTGACAGAA-3') were generated by Golden Gate cloning, as previously described (9). To generate pAdShu.TALEN-L[S1].IRES.RFP and pAdShu.TALEN-R[S1].IRES.RFP, plasmids 1383.pVAX.AAVS1.TALEN.L-94 and 1384.pVAX.AAVS1.TALEN.R-95 were digested with *Eco*32I (Fermentas) and *Pme*I (New England Biolabs) and the resulting 3.0-kb *TALEN* open reading frame (ORF) fragment were isolated and ligated into the *Eco*32I-linearized 8.8-kb plasmid vector backbone pShuttle. IRES.DsRedEx2.1. Cells transduced with vectors based on this construct can be traced and quantified via the encoded red fluorescent protein reporter DsRedEx2.1 (RFP). The firstgeneration (i.e. early region 1 [*E1*]-deleted) fiber-modified adenoviral vector molecular clones pAd.ΔE1.TALEN-L[S1].F[50] and pAd.ΔE1.TALEN-L[S1].F[50] (Figure 2A) were assembled by HR following transformation of E. coli cells BJ5183[pAdEasy-1.50] (23) with shuttle plasmids pAdShu. TALEN-L[S1].IRES.RFP and pAdShu.TALEN-R[S1].IRES.RFP, respectively, whilst the second-generation adenoviral vector molecular clones pAd.ΔE1ΔE2A.TALEN-L[S1].F[50] and pAd. ΔE1ΔE2A.TALEN-R[S1].F[50] (Figure 2A) were assembled by HR following transformation of *E. coli* cells BJ5183[pAdEasy-2.50] (23) with shuttle plasmids pAdShu.TALEN-L[S1].IRES.RFP and pAdShu. TALEN-R[S1].IRES.RFP, respectively. Prior to transformation all shuttle plasmids were digested to completion with *Mss*I. The structure and composition of lentiviral vector constructs pLV. TALEN-L[S1].i.RFP, pLV.TALEN-R[S1].i.RFP, pLV.TALEN-L[S1] and pLV.TALEN-R[S1] are described in Figure 1A. The plasmid CMVPRES (24; herein referred to as pLV.CMV.eGFP) served as a negative control in the PCR amplifications of *TALEN* sequences.

Viral Vectors

The production of adenoviral vectors Ad.ΔE1.TALEN-L[S1].F[50] and Ad.ΔE1.TALEN-R[S1].F[50] took place in PER.C6 cells, whereas that of adenoviral vectors Ad.ΔE1ΔE2A.TALEN-L[S1].F[50] and Ad.ΔE1ΔE2A.TALEN-R[S1].F[50] occurred in PER.E2A cells. The rescue of vector particles was initiated by transfecting each of the *Pac*I-linearized adenoviral vector molecular clones into their respective packaging cells by using a polyethyleneimine-based protocol (23). The procedures to propagate and purify the resulting viral vectors have been described elsewhere (23). Vector titers were determined through limiting dilutions on HeLa indicator cells seeded at a density of 0.8×10^5 cells per well of 24-well plates (Greiner). At three days post-transduction, frequencies of RFP-positive cells were determined through flow cytometry by using a BD LSR II FACS (Becton Dickinson). As a result, the titers are expressed in terms of HeLa cell-transducing units (HTUs) per ml. The HIV-1-based and vesicular stomatitis virus

glycoprotein-G (VSV-G)-pseudotyped lentiviral vectors were generated by transfecting 293T cells essentially as specified elsewhere (25) except for the use of 7.5×10^6 cells in T75-cm^2 flasks instead of 1.75×10^7 cells in T175-cm^2 flasks. A total amount of 12.85 µg of DNA was transfected corresponding to 2:1:1 mixtures of lentiviral vector transfer plasmid, packaging construct psPAX2 (Addgene plasmid 12260) or psPAX2^{D116N} (25) and VSV-G pseudotyping construct pLP/VSVG (Invitrogen). Lentiviral vector titers were determined by using HIV-1 p24^{gag} immunocapture (Perkin Elmer).

Immunofluorescence microscopy

Fifty-thousand HeLa cells were seeded in wells of 24-well plates and, after an overnight incubation period, they were either mock-transduced or were transduced with Ad.ΔE1. TALEN-L^{S1}.F^{50} or with Ad.ΔE1.TALEN-R^{S1}.F^{50} at 3 HTU/cell. Three days post-transduction, the cells were subjected to immunostaining essentially as described elsewhere (26) except for the use of an influenza hemagglutinin (HA)-specific mouse monoclonal IgG1 antibody (HA.11 clone 16B12; Covance) together with a goat anti-mouse IgG (H+L) Alexa 488-conjugated secondary antibody (A-11001; Molecular Probes).

Western blot analysis

HeLa cells seeded at a density of 1×10^5 cells per well of 24-well plates (Greiner) were transduced with different amounts of Ad.ΔE1.TALEN-L^{S1}.F^{50} or Ad.ΔE1.TALEN-R^{S1}.F^{50}. At 72 hours post-transduction, the cell monolayers were rinsed with ice-cold phosphate-buffered saline (pH 7.4) and lysed in 100 µl of ice-cold RIPA buffer composed of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1% sodium dodecyl sulphate (SDS), 0.5% sodium deoxycholate and 1% Nonidet P-40 plus a cocktail of protease inhibitors (Complete Mini, Roche Applied Science). Next, the cell lysates were spun at 4°C for 8 min at 16,400×g. The proteins were subjected to electrophoresis through an SDS-7.5% polyacrylamide gel after which they were transferred onto an Immobilon-P membrane (Millipore) by electroblotting overnight. After incubation with TBST blocking solution consisting of 10 mM Tris-HCl (pH 8.0), 150 mM NaCl and 0.2% Tween-20 supplemented with 10% non-fat dry milk powder (Elk, Campina), the membranes were incubated overnight at 4°C with the HA-specific monoclonal antibody F-7 (SC-7392, Santa Cruz Biotechnology) diluted 1:1000. Subsequently, the membranes were washed once with TBST at room temperature and subjected to a 3-hour incubation period with a horseradish peroxidase conjugated goat anti-mouse secondary antibody (SC-2005; Santa Cruz Biotechnology) diluted 1:5000 in blocking solution. Finally, the membranes were rinsed and processed for protein detection by chemiluminescence as described elsewhere (27).

The Western blot analysis to detect *TALEN* expression following lentiviral vector transduction was performed as follows. Hundred-thousand HeLa cells were seeded in wells of 6- well plates and were transduced with 200 ng of p24^{gag} of LV.TALEN-L^{S1} or 400 ng of p24^{gag}

of IDLV.TALEN-L[S1] in the presence of polybrene (8 μg/ml). Forty-eight hours after transduction, the cells were harvested and total protein lysates were prepared as aforementioned. Fifty micrograms of protein were subjected to electrophoresis through a NuPAGE 4-12% Bis-Tris Gel (Invitrogen) and were transferred onto a nitrocellulose membrane (Hybond-ECL, Amersham) that was subsequently exposed overnight at 4°C to a mouse monoclonal antibody against the HA-tag (clone 12CA5 a gift from Valeria Marigo). A horseradish peroxidase-conjugated goat anti-mouse secondary antibody (Amersham) was diluted 1:2000 in blocking solution. The proteins were detected by using a chemiluminescence labelling detection reagent (ECL; GE Healthcare Life Sciences). For protein loading normalization, the same membrane was stripped and incubated with a 2000-fold diluted mouse monoclonal antibody directed against β-actin (sc-81178, Santa Cruz Biotechnology).

Isolation and structural analysis of adenoviral vector DNA
DNA was isolated from purified adenoviral vector particles as follows. To a 50-μl vector preparation aliquot it was added 12 μl of DNaseI buffer (130 mM Tris-HCl [pH 7.5], 1.2 mM CaCl$_2$ and 50 mM MgCl$_2$), 8 μl of DNaseI (10 mg/ml; Fermentas) and 50 μl of Milli-Q water (Millipore). This mixture was incubated for 30 min at 37°C, after which, the DNaseI was inactivated for 1 hour at 55°C after the addition of 2.4 μl of 0.5M EDTA (pH 8.0), 6 μl of 10% SDS and 1.5 μl of a Proteinase K (21.4 mg/ml; Fermentas). Next, the vector DNA was purified by using the QIAEX II Gel Extraction Kit (Qiagen) and was eluted in 30 μl of 10 mM Tris-HCl (pH 7.8). After measuring its concentration with the aid of a Nanodrop (Thermo Scientific), the recovered DNA was subjected to *Bcu*I (Fermentas) digestion and agarose gel electrophoreses (200 ng) and to PCR analyses (1 ng). The PCR amplifications were carried out with primers TALEN-Seq-F (5'-CCAGCTGCTGAAGATCGCCAAGC-3') and TALEN-Seq-R (5'-CTATCCTGAGTGGAATTTCTGGC-3') whose target sequences flank the TALEN DNAbinding motifs. The DNA samples were subjected to PCR with 0.4 μM of these primers (Eurofins MWG Operon), 0.1 mM of each dNTP (New England Biolabs), 1× Colorless GoTaq Flexi buffer, 1 mM MgCl$_2$ and 2.5 units of GoTaq DNA polymerase (all from Promega). The PCR cycles were performed in a DNA Engine Tetrad 2 Peltier Thermal Cycler (Bio-Rad) using the following cycling parameters. After a 3-min denaturation step at 95°C, the samples were subjected to 30 cycles consisting of 30 s at 95°C, 30 s at 60°C and 4 min at 72°C. The reactions were terminated by a final elongation period of 3 min at 72°C. One-fifth of these reactions were subjected to electrophoresis through 1% agarose gels in 1× TAE buffer. Moreover, to establish the integrity of *TALEN* sequences in adenoviral vector particles at the nucleotide level, the 2.4-kb *Bcu*I fragments from Ad.ΔE1.TALEN-L[S1].F[50] or from Ad.ΔE1.TALEN-R[S1].F[50] DNA were inserted into a *Xba*I-linearized pUC19 backbone and were sequenced by using M13-F (5'-GACGTTGTAAAACGACGGCCAGT-3') and M13-R (5'-CAGGAAACAGCTATGACCATGA-3') primers.

Surveyor nuclease S mutation detection assay

Genomic DNA was extracted from mock-transduced or adenoviral vector-transduced target cells as described before (28). Briefly, at 72 hours post-infection, the cells were collected by trypsinization and incubated overnight at 55°C in 500 µl of lysis buffer (100 mM Tris-HCl [pH 8.5], 5 mM EDTA, 0.2% SDS, 200 mM NaCl) supplemented with freshly-added Proteinase K at a final concentration of 100 ng/ml. Next, the cell lysates were extracted twice with a buffersaturated phenol:chloroform:isoamyl alcohol mixture (25:24:1) and once with chloroform. The DNA in the aqueous phase was precipitated by the addition of 2.5 volumes of absolute ethanol and 0.5 volumes of 7.5 M ammonium acetate (pH 5.5). After washing with 70% ethanol, the DNA pellet was mildly dried and dissolved in 100 µl of TE buffer (10 mM Tris-HCl [pH 8.0] and 1mM EDTA) supplemented with RNase A (Fermentas) at a final concentration of 100 µg/ml.

The recovered DNA was used for PCR analyses. The extent of *AAVS1* processing by the error-prone NHEJ DNA repair pathway was evaluated by using the Surveyor assay (SURVEYOR Mutation Detection Kit, Transgenomics) according to the manufacturer's instructions. In brief, genomic DNA purified from mock-transduced or adenoviral vectortransduced cells was used as a template to PCR amplify a 469-bp *AAVS1* sequence flanking the TALEN pair target site. PCR mixtures contained 0.4 µM of primers AAVS1-CEL1-F (5'-TTCGGGTCACCTCTCACTCC-3') and AAVS1-CEL1-R (5'-GGCTCCATCGTAAGCAAACC-3'), 0.1 mM of each dNTP, 1× Colorless GoTaq Flexi buffer, 1mM $MgCl_2$ and 2.5 units of GoTaq DNA polymerase. The PCR cycles were performed using the following cycling conditions. An initial denaturation step at 95°C for 5 min was followed by 40 cycles of 30 s at 95°C, 30 s at 61°C and 30 s at 72°C. The reactions were terminated by a final elongation period of 5 min at 72°C. The PCR products, after being heat-denatured and re-annealed according to the manufacturer's specifications, were treated with the nucleotide mismatch-sensitive Surveyor nuclease S (SNS) enzyme in order to detect small insertions and deletions (indels) caused by NHEJ-dependent DNA repair. Samples were loaded on a 1.7% agarose gel and electrophoresis was performed in 1× TAE buffer. The DNA was visualized by in-gel staining with 2 µg/ml of ethidium bromide (Sigma-Aldrich) in 1× TAE buffer followed by stain removal in Milli-Q water. The frequencies of target DNA cleavage were determined by using the ImageJ software (National Institutes of Health, USA). Genomic DNA derived from hMSC cells co-transduced with a 1:1 mixture of Ad.ΔE1.TALEN-L[S1].F[50] and Ad.ΔE1.TALEN-R[S1].F[50] at a total multiplicity of infection (MOI) of 24 HTU/cell, served as template to PCR amplify the *AAVS1* target site region. The pGEM-T Easy TA cloning vector (Promega) was ligated to the resulting 469-bp amplicons and was transformed into *E. coli* strain DH5α. Plasmid DNA from 20 randomly selected colonies was extracted and the respective inserts were subjected to DNA sequencing by using the M13-F primer.

PCR analyses of TALEN sequences following lentiviral vector-mediated gene transfer

Total cellular DNA from HeLa cells incubated with integration-competent or integrase-defective lentiviral vectors (IDLVs) was isolated as specified elsewhere (28). The recovered DNA was subjected to three different PCR protocols. To determine the integrity of the cellular DNA, a PCR targeting the human housekeeping gene hypoxanthine phosphoribosyltransferase 1 (*HPRT1*) was performed. To this end, reactions containing 0.4 µM of primers hHPRT.1 and hHPRT.2 (29), 0.1 mM of each dNTP, 1× Colorless GoTaq Flexi buffer, 1mM MgCl$_2$ and 2.5 U of GoTaq DNA polymerase, were assembled. The PCR cycles were performed using the following cycling conditions. An initial denaturating step at 95°C for 5 min was followed by 30 cycles of 30 s at 95°C, 30 s at 60°C and 120 s at 72°C.

A second set of PCR mixtures were assembled to amplify a 458-bp *TALEN* sequence corresponding to the N-terminal, non-repetitive region, of each nuclease. These consisted of 0.4 µM of primers TAL-N.s (5'-CCTTACGACGTGCCTGACTACG-3') and TAL-N.as (5'-CGCTTGGCGATCTTCAGC-3'), 0.1 mM of each dNTP, 1× Colorless GoTaq Flexi buffer, 1mM MgCl$_2$ and 2.5 U of GoTaq DNA polymerase. The PCR cycles were performed using the following cycling conditions. An initial denaturating step at 95°C for 5 min was followed by 30 cycles of 30 s at 95°C, 30 s at 62°C and 30 s at 72°C.

Finally, the third PCR set was designed to amplify a 2.5-kb *TALEN* sequence encompassing the region encoding the repetitive DNA-binding domain of each nuclease. To this end, reactions containing 0.4 µM of primers CMV-prom.F (5'-AATGGGCGGTAGGCGTGTA-3') and TALE.as.Pr2 (5'-GTCGGGTCTGCTCAGCTG-3'), 0.1 mM of each dNTP, 1× Colorless GoTaq Flexi buffer, 1mM MgCl$_2$ and 2.5 U of GoTaq DNA polymerase, were assembled. The PCR cycling parameters consisted of an initial denaturating step at 95°C for 5 min followed by 40 cycles of 30 s at 95°C, 30 s at 60°C and 180 s at 72°C. In all three protocols, the PCR amplifications were terminated by a final elongation period of 5 min at 72°C. Control samples subjected to the same PCR conditions were carried out in parallel. These controls consisted of using, in place of DNA from lentiviral vector-transduced HeLa cells, DNA from mocktransduced HeLa cells, DNA from Ad.ΔE1.TALEN-R[S1].F[50] -transduced HeLa cells, pAd.ΔE1.TALEN-L[S1].F[50], pLV.CMV.eGFP or MilliQ water. Additional controls were provided by PCR amplifications on DNA isolated from HeLa cells exposed to an LV.TALEN-R[S1] preparation made with the aid of pUC19 instead of the lentiviral vector packaging construct psPAX2 (Addgene plasmid 12260) or on DNA extracted from cells transfected with pLV.TALEN-R[S1] transfer plasmid mixed or not mixed with ExGen500 (Fermentas).

Southern blot analysis

Total cellular DNA from 4×10^5 hMSCs exposed to three different MOIs of Ad.ΔE1.TALEN-L[S1].F[50] or of Ad.ΔE1.TALEN-R[S1].F[50] was extracted 72 hours post-transduction and purified as described before (28). After overnight digestion with *Bam*HI (Fermentas), 10 µg of

DNA per sample was resolved in a 1% agarose gel in 1× TAE buffer. Next, the DNA was transferred by capillary action onto an Amersham Hybond-XL membrane (GE Healthcare Life Sciences) using a standard Southern blot technique. A 397-bp probe specific for *TALEN* DNA corresponding to the N-terminal region of the nucleases was isolated by digestion of 1383. pVAX.AAVS1.TALEN.L-94 with *Xba*I and *Pst*I (both from Fermentas) and was purified with the aid of the QIAEX II gel extraction kit (Qiagen). This DNA probe was radiolabelled with EasyTide (α-^{32}P) dCTP (3000 Ci/mmol; Perkin Elmer) by using the DecaLabel DNA labelling system (Fermentas). Prior to its addition onto the membrane, the radiolabeled fragment was separated from unincorporated dNTPs through size-exclusion chromatography with the aid of Sephadex-50 columns (GE Healthcare Life Sciences). A Storm 820 PhosphorImager (Amersham) as used for the detection of the probe-hybridizing DNA. Finally, images were acquired and processed by using the Storm scanner control 5.03 and the ImageQuant Tools 3.0 software, respectively (both from Amersham).

HeLa cell clones were derived from an LV.TALEN-L[S1]-transduced population by limiting dilution using a seeding density of 0.3 cells per well in 96-well plates. Southern blot analysis of genomic DNA extracted from various HeLa cell clones was performed as follows. Genomic DNA was extracted from $5×10^6$ cells from each clone using a QIAamp DNA Mini Kit according to the manufacturer's instructions (Qiagen). Ten micrograms of genomic DNA was digested overnight with *Afl*II (New England Biolabs), subjected to electrophoresis through a 0.8% agarose gel and was transferred onto a Hybond-XL nylon membrane. Prior to DNA detection, this membrane was incubated with a (α-^{32}P)-labelled 814-bp probe corresponding to the lentiviral vector LTR elements R and U5, the packaging signal and part of the Rev-responsive element (RRE).

RESULTS

Lentiviral vectors introduce rearranged *TALEN* genes into human cells

We started by investigating the feasibility of generating lentiviral vectors to deliver TALENencoding expression units into human cells. Although the long-term presence of artificial nucleases in cells is, clearly, undesirable in most experimental settings, lentiviral vectors can be made with in-built stringent drug-controllable transgene expression systems (30) or class I integrase mutant moieties (31,32). These integrase mutations are non-pleiotropic as they confer an integration-defective phenotype to the vectors by specifically disabling the chromosomal insertion step of their DNA. IDLVs are, therefore, being increasingly exploited to deliver transgenes whose expression is only required during a short timeframe in populations of cycling cells. Examples include IDLV-mediated delivery of transposases, homing endonucleases and zinc finger nucleases (33-36).

Therefore, we constructed the transfer plasmids pLV.TALEN-L^{S1} and pLV.TALEN-R^{S1} to package the lentiviral vector pairs LV.TALEN-L^{S1}/IDLV.TALEN-L^{S1} and LV.TALEN-R^{S1}/IDLV.TALEN-R^{S1}, respectively. These constructs encode a TALEN enzyme (i.e., TALEN-L^{S1} or TALEN-R^{S1}) specific for the *AAVS1* locus at the human chromosome 19 (19q13.42-qter). Moreover, we also constructed pLV.TALEN-L^{S1}.i.RFP and pLV.TALEN-R^{S1}.i.RFP to make lentiviral vector particles carrying bicistronic cassettes encoding TALEN-L^{S1} and TALEN-R^{S1}, respectively, and through an IRES, the RFP reporter. To investigate the integrity of vector genomes in target cells we carried out transduction experiments on HeLa cells deploying the lentiviral vector set harboring the *TALEN-R^{S1}* ORF. Following DNA extraction from vectortransduced cells, we set-up a PCR assay (Figure 1A) based on the use of primer pairs framing the repeats of the *TALEN* ORF and its upstream region encoding the nuclease's N-terminus (CMVprom. F/TALE.as.Pr2) or targeting exclusively the latter segment (TAL-N.s/TAL-N.as). PCR amplifications with primers CMV-prom.F and TALE.as.Pr2 on DNA extracted from HeLa cells transduced with LV.TALEN-R^{S1}, IDLV.TALEN-R^{S1} or LV.TALEN-R^{S1}.i.RFP did not yield detectable levels of 2.5-kb amplicons diagnostic for full-length *TALEN-R^{S1}* sequences (Figure 1B, upper panels, lanes 5, 6 and 11, respectively). In fact, in these samples, the only PCR products that could be discerned were those consistent with heterogeneously sized sub-genomic templates whose molecular-weight range spanned from about 0.8 to 1.5 kb. Significantly, PCR amplifications carried out on the same DNA templates with primers TAL-N.s and TAL-N.as, gave rise to properly-sized 0.46 kb PCR products indicating that the non-repetitive upstream portion of the *TALEN-R^{S1}* ORF did not suffer major rearrangements in transduced cells (Figure 4B, middle panels). Of note, equivalent results were obtained in complementary transduction experiments that made use of the lentiviral vector set carrying the *TALEN-L^{S1}* ORF (not shown). To validate these data, PCR amplifications were done in parallel on control samples. These negative and positive controls consisted of water (Figure 1B, lane 4) or DNA extracted from mock-transduced HeLa cells (Figure 1B, lane 1), Ad.ΔE1.TALEN-R^{S1}.F^{50}-transduced HeLa cells (Figure 1B, lane 10), plasmid pAd.ΔE1.TALEN-L^{S1}.F^{50} (Figure 1B, lane 2) or plasmid pLV.CMV.eGFP (Figure 1B, lane 3). Additional controls consisted of PCR amplifications on DNA isolated from HeLa cells exposed to an LV.TALEN-R^{S1} preparation made with pUC19 instead of the lentiviral vector packaging construct psPAX2 (Figure 1B, lane 7) or incubated with plasmid pLV.TALEN-R^{S1} mixed or not mixed with the DNA transfection agent ExGen500 (Figure 1B, lanes 8 and 9, respectively). PCR amplifications targeting *HPRT1* served as a control for DNA template integrity (Figure 1B, lower panels).

To complement these PCR results on whole target cell populations, we carried out Southern blot analysis on genomic DNA from individual HeLa cell clones derived from a bulk population transduced with LV.TALEN-L^{S1}. To this end, the chromosomal DNA from 15 clones was digested with AflII, Southern blotted and incubated with the probe depicted in Figure 1A. To serve as a molecular-weight reference, the parental transfer plasmid pLV.TALEN-L^{S1}

was subjected to the same procedures. Results depicted in Figure 1C demonstrate that none of the tested clones yielded the *Afl*II 6.4-kb fragment diagnostic for full-length LV.TALEN-L[S1] proviruses.

Finally, we investigated through HA tag-directed Western blot analysis, *TALEN-L[S1]* expression following addition of 200 ng of p24[gag] of LV.TALEN-L[S1] or 400 ng of IDLV.TALEN-L[S1] onto HeLa cultures. These experiments failed to show detectable levels of HA-specific signals in protein lysates corresponding to vector-transduced cells, whilst HA-specific signals at a molecular-weight consistent with full-length TALEN-L[S1] proteins could be readily detected in control lysates derived from pLV.TALEN-L[S1]-transfected HeLa cells or 293T producer cells (Figure 1D). To serve as controls for assembly of LV.TALEN-L[S1] or IDLV.TALEN-L[S1] particles, we produced in parallel LV and IDLV vector preparations carrying a human *PGK1* promoter-driven eGFP reporter (LV.PGK.eGFP and IDLV.PGK.eGFP, respectively). Exposure of HeLa cell cultures to 200 ng of p24[gag] of LV.PGK.eGFP or to 400 ng of IDLV.PGK.eGFP led to the transduction of 97% and 83% of the target cells, respectively (not shown).

Taken together, these data indicate that *TALEN* sequences introduced into human cells through lentiviral vectors are prone to extensive rearrangements resulting presumably from deletions involving the *TALE* repeats.
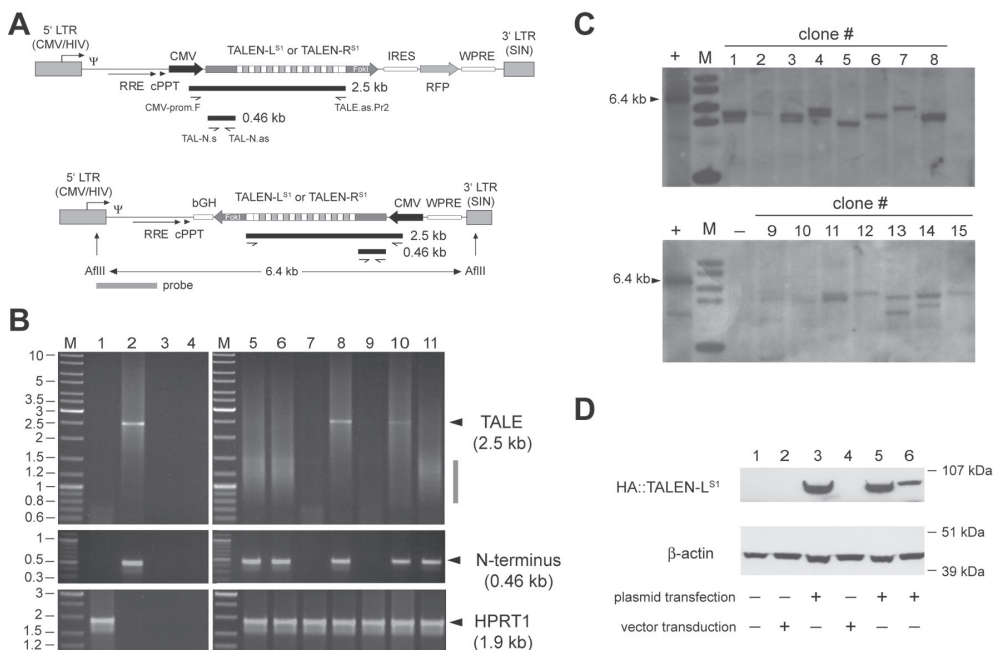
Figure 1. Structural analyses of lentiviral vectors harboring *TALEN* sequences in transduced cells. **(A)** Genetic organization of lentiviral vector constructs. Diagram of bicistronic pLV.TALEN-L^{S1}.i.RFP and pLV.TALEN-R^{S1}.i.RFP and of monocistronic pLV.TALEN-L^{S1} and pLV.TALEN-R^{S1} lentiviral vector transfer plasmids. The bicistronic constructs code for the *AAVS1*-specific custom-made nucleases TALEN-L^{S1} or TALEN-R^{S1} and, through an IRES sequence, a RFP reporter, whereas the monocistronic plasmids encode exclusively the *AAVS1*-specific designer nucleases. Each of the TALEN proteins are tagged by an HA antigen located close to their N-termini (not drawn). Gray boxes with broken arrow, hybrid 5' long-terminal repeat (LTR) containing CMV and HIV-1 sequences; gray boxes without broken arrow, selfinactivating (SIN) 3' LTR; Ψ, HIV-1 packaging signal; RRE, Rev-responsive element; cPPT, central polypurine tract; WPRE, Woodchuck hepatitis virus posttranscriptional regulatory element. In pLV. TALEN-L^{S1}.i.RFP and pLV.TALEN-R^{S1}.i.RFP the ORFs are under the transcriptional control of the CMV promoter and the polyadenylation signal present within the SIN 3' LTR, whilst the ORFs in pLV.TALEN-L^{S1} and pLV.TALEN-R^{S1} are under the control of the CMV promoter and the bovine *growth hormone* polyadenylation signal (bGH). Primers, PCR products and Southern blot probe used are depicted as half-arrows, black bars and gray bar, respectively, and are drawn in relation to their respective vector DNA sequences. For the sake of simplicity, plasmid backbone sequences are not show. The sizes of the PCR products expected from the amplification of full-length lentiviral vector genomes are indicated. The 6.4-kb *Afl*II restriction fragment corresponding to intact DNA from LV.TALEN-L^{S1} or LV.TALEN-R^{S1} is also indicated. **(B)** Expression analysis. PCR analyses of TALEN-encoding lentiviral vector genomes following transduction of HeLa cells. PCR amplifications with the aid of the primers shown in Figure 1A were carried out using, as template, total cellular DNA from mocktransduced HeLa cells, plasmid pAd. ΔE1.TALEN-L^{S1}.F^{50}, plasmid pLV.CMV.eGFP or nuclease-free H$_2$0 (lanes 1, 2, 3 and 4, respectively). PCR amplifications using the same primer pairs were performed in parallel on DNA extracted from HeLa cells transduced with integrationcompetent LV.TALEN-R^{S1} (lane 5), integrase-defective IDLV.TALEN-R^{S1} (lane 6), Ad.ΔE1.TALEN-R^{S1}.F^{50} (lane 10) or integration-competent LV.TALEN-R^{S1}.i.RFP (lane 11). Extra PCR controls supplementing those corresponding to lanes 1 through 4, were provided by using DNA isolated from HeLa cells exposed to an LV.TALEN-R^{S1} preparation made in parallel but deploying pUC19 instead of lentiviral vector packaging construct psPAX2 (lane 7) or incubated with pLV.TALEN-R^{S1} mixed or not mixed with the DNA transfection agent ExGen500 (lanes 8 and 9, respectively). Lane M, Gene Ruler DNA Ladder Mix. A primer pair targeting the human *HPRT1* served to control the integrity of the cellular DNA. **(C)** Clonal analysis. Southern blot analysis of *Afl*II-digested genomic DNA from HeLa cell clones 1 through 15 stably transduced with LV.TALEN-L^{S1}. The vector-specific probe that was used is shown in Figure 1A (horizontal gray bar). Lane M, 1 Kb molecular-weight marker. The parental transfer plasmid pLV.TALEN-L^{S1} treated with *Afl*II served as an internal control for intact vector DNA. **(D)** TALEN protein detection. HA tag-directed Western blot analysis of full-length *TALEN-LS1* expression in 293T producer cells (lane 1), IDLV.TALEN-L^{S1}- and LV.TALEN-L^{S1}-transduced HeLa cells (lane 2 and lane 4, respectively) and in pLV.TALEN-L^{S1}-transfected 293T cells used to generate these IDLV.TALEN-L^{S1} and LV.TALEN-L^{S1} vector preparations (lanes 3 and 5, respectively). Lane 6, HeLa cells transfected with pLV. TALEN-L^{S1}. The β-actin served as loading control.
◄

Adenoviral vectors deliver full-length TALEN genes into human cells inducing high-level DSB formation at an endogenous chromosomal target locus

In light of the previous results, we sought to investigate the feasibility of using adenoviral vectors as an alternative delivery platform for the introduction of intact *TALEN* genes into human cells. To this end, the full-length adenoviral vector molecular clones pAd.ΔE1.TALEN-L^{S1}.F^{50} and pAd.ΔE1.TALEN-R^{S1}.F^{50} were assembled through HR in *E. coli* to produce the *E1*-deleted adenoviral vectors Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50}, respectively (Figure 2A). These clones contain, in place of *E1* from the prototypic adenovirus serotype
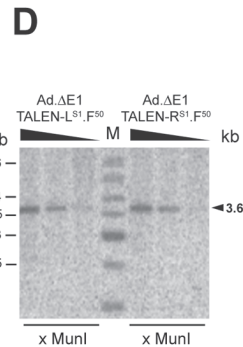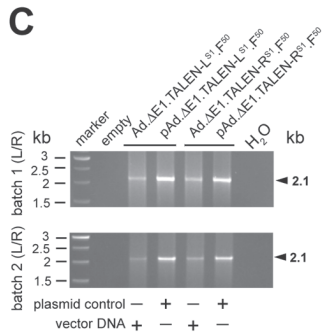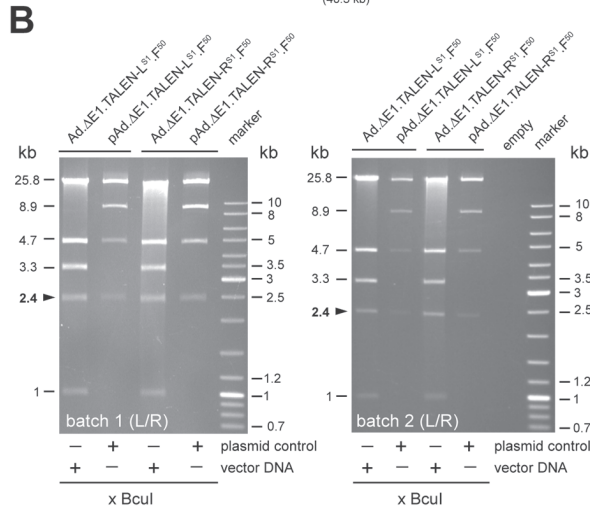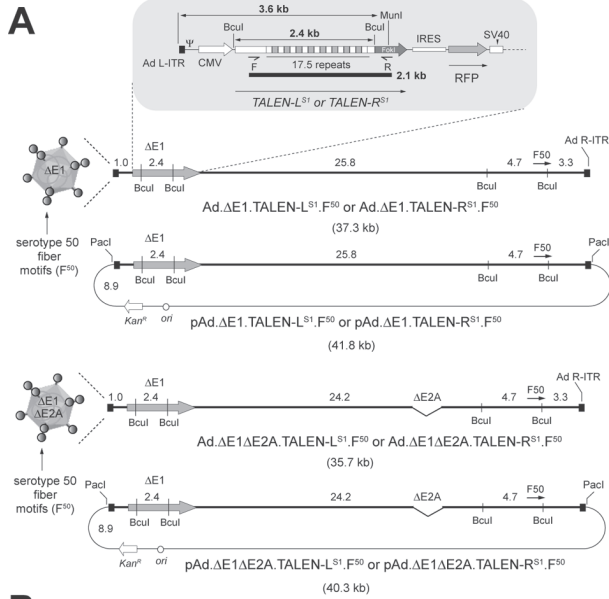
5, a bicistronic expression unit encoding a monomer of the *AAVS1*-specific TALEN pair and, for tracing and quantification purposes, the RFP reporter (Figure 2A). In addition, their adenoviral genomic sequences harbor an ORF coding for a vector retargeting chimeric fiber composed of serotype 5 basal shaft domains fused to apical shaft and knob domains derived from the species B adenovirus serotype 50 (F^{50}). The latter CD46-interacting motif allows for entry of vectors displaying species B fiber domains into human cells lacking on their surface the Coxsackie B virus and adenovirus receptor, CAR. Of note, among these cells there are various therapeutically relevant tissue-specific normal and malignant stem and progenitor cell types (37-41). In addition, plasmids pAd.ΔE1ΔE2A.TALEN-L^{S1}.F^{50} and pAd.ΔE1ΔE2A.TALEN-R^{S1}.F^{50} were also constructed to assemble fiber-modified *E1*- plus *E2A*-deleted particles Ad.ΔE1ΔE2A.TALEN-L^{S1}. F^{50} and Ad.ΔE1ΔE2A.TALEN-R^{S1}.F^{50}, respectively (Figure 2A). By virtue of their lack of E2A, these so-called second-generation adenoviral vectors have been shown to display a muchreduced *de novo* replication and "leaky" viral gene expression profile in target cells when compared to their *E1*-deleted, first-generation, counterparts (see, for instance, ref. 20).

We started by testing whether *TALEN* sequences are stably maintained in first- and second-generation adenoviral vectors following their rescue and subsequent propagation on their respective packaging cell lines. Thus, these producer cells were transfected with *Pac*I-treated plasmids pAd.ΔE1.TALEN-L^{S1}.F^{50}, pAd.ΔE1.TALEN-R^{S1}.F^{50}, pAd.ΔE1ΔE2A.TALEN-L^{S1}.F^{50} or pAd.ΔE1ΔE2A.TALEN-R^{S1}.F^{50} to assemble the corresponding adenoviral vector particles (Figure 2A). The rescued particles were subsequently amplified following serial propagation on increasing amounts of packaging cells. The frequency of RFP-positive producer cells, as monitored by direct fluorescence microscopy, revealed a clear propagation round-dependent increase in vector yields (not shown). The Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} vectors, corresponding to two independent batches, were purified through CsCl buoyant density gradient ultracentrifugation, after which, the integrity of their genomes was probed by BcuI restriction fragment length analysis. To serve as an additional molecular-weight reference, the parental plasmids pAd.ΔE1.TALEN-L^{S1}.F^{50} and pAd.ΔE1.TALEN-R^{S1}.F^{50} were taken along in this analysis. Results depicted in Figure 2B show that the overall DNA restriction pattern is consistent with that expected from linear full-length adenoviral vector genomes (Figure 2A). Thus, importantly, the internal *TALE*-encompassing restriction fragments derived from Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} DNA did co-migrate with those corresponding to their respective 2.4-kb parental plasmid counterparts (Figure 2B, arrowheads). Results equivalent to these were obtained following the structural analyses of DNA isolated from Ad.ΔE1ΔE2A.TALEN-L^{S1}.F^{50} and Ad.ΔE1ΔE2A. TALEN-R^{S1}.F^{50} (Supplementary Figure S2). Next, we performed a more sensitive PCR-based assay using primers bracketing the 17.5 repeats of *TALEN-L^{S1}* and *TALEN-R^{S1}* to search for rearranged *TALEN* sequences in the purified vector preparations. Again, parental constructs

pAd.ΔE1.TALEN-L^{S1}.F^{50} and pAd.ΔE1.TALEN-R^{S1}.F^{50} served as control DNA templates. This assay led to the detection of 2.1-kb amplicons whose size is diagnostic for vector DNA genomes harboring intact TALEN-encoding genes (Figure 2C). Moreover, Southern blot analysis of extrachromosomal DNA isolated from hMSCs exposed to different amounts of Ad.ΔE1.TALEN-L^{S1}.F^{50} or Ad.ΔE1.TALEN-R^{S1}.F^{50} also led to the sole detection of full-length *TALEN* sequences in the transduced cells (Figure 2D). Finally, we cloned into pUC19 the 2.4-kb DNA species from *Bcu*I-treated Ad.ΔE1.TALEN-L^{S1}.F^{50} or *Bcu*I-treated Ad.ΔE1.TALEN-R^{S1}.F^{50} and sequenced the inserts from six individual clones. DNA sequence information corresponding to the 7 most downstream and to the 3 to 4 most upstream *TALEN* repeats could be retrieved (Supplementary Figure S3). The nucleotide sequences analyzed revealed to be isogenic to that in the parental constructs with no evidence for point mutations, local rearrangements and/or deletions.

Transduction experiments deploying Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} alone or mixed together were carried out on CAR-positive human cervix carcinoma HeLa cells (42) and on two CAR-negative cell types, namely, immortalized myoblasts from a Duchenne muscular dystrophy (DMD) patient and hMSCs (42,43). The use of relatively modest MOIs was sufficient to transduce most target cells in the various cultures (Figure 3A). Importantly, HA tag-specific immunofluorescence microscopy and Western blot analyses (Figure 3B and 3C, respectively) demonstrated the expression of *TALEN-L^{S1}* and *TALEN-R^{S1}* genes in target HeLa cells following their transduction with Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}. F^{50}, respectively. Taken together, we conclude that first-generation adenoviral vectors can deliver expression units encoding full-length TALENs into human cells.

3

Figure 2. Structural analysis of TALEN-encoding adenoviral vector genomes. **(A)** DNA structures of the first-generation adenoviral vectors Ad.ΔE1.TALEN-L[S1].F[50] and Ad.ΔE1.TALEN-R[S1].F[50] and of the second-generation adenoviral vectors Ad.ΔE1ΔE2A.TALEN-L[S1].F[50] and Ad.ΔE1ΔE2A.TALEN-R[S1].F[50] drawn in relation to those of their respective plasmid clones. The adenoviral vector and shuttle plasmid DNA templates are depicted as *Bcu*I physical maps with numerals specifying the different restriction fragment sizes in kilobases (kb). These DNA molecules harbor a bi-cistronic expression unit coding for the *AAVS1*-specific artificial nucleases TALEN-L[S1] or TALEN-R[S1] and, through an IRES sequence, a RFP reporter. The open reading frames are under the transcriptional control of the cytomegalovirus immediate-early gene promoter (CMV) and the simian virus 40 (SV40) polyadenylation signal. The central region of *TALEN-L[S1]* and *TALEN-R[S1]* encode a tandem of 17.5 TALE repeats that define specificity to the *AAVS1* target sequence at the human chromosome 19 (19q13.42-qter). C-terminally both TALENs contain the non-specific cleavage domain of FokI, whose catalytic activity is contingent upon DNA binding-dependent dimerization. Each of the TALEN proteins are tagged by an HA antigen located close to their Ntermini (not drawn). The thick and thin lines represent adenoviral vector and plasmid backbone sequences, respectively. Ψ, human adenovirus serotype 5 packaging signal; Ad L-ITR and Ad RITR, "left" and "right", respectively, adenoviral inverted terminal repeat. F50, ORF encoding chimeric fiber composed of basal shaft domains from adenovirus serotype 5 fused to apical shaft and knob motifs from adenovirus serotype 50. The kanamycin-resistance gene (*Kan[R]*) and the prokaryotic origin of replication (*ori*) are also indicated. **(B)** Restriction fragment length analysis of Ad.ΔE1.TALEN-L[S1].F[50] and Ad.ΔE1.TALEN-R[S1].F[50] genomes treated with *Bcu*I. Plasmids pAd.ΔE1.TALEN-L[S1].F[50] and Ad.ΔE1.TALEN-R[S1].F[50] served as a reference. The sizes (in kb) of the DNA restriction fragments resulting from the BcuI treatment are indicated on the left. The arrowhead points to the 2.4 kb DNA fragments diagnostic for intact, non-rearranged, *TALE*-derived DNA sequences within TALEN-encoding genes. Marker, molecular-weight marker Gene Ruler DNA Ladder Mix. **(C)** PCR analysis on DNA isolated from Ad.ΔE1.TALEN-L[S1].F[50] and Ad.ΔE1.TALEN-R[S1].F[50] particles using primers TALEN-Seq-F and TALEN-Seq-R (F and R half arrows, respectively). These primers, drawn in relation to their respective target sequences, amplify most of the *TALEN-L[S1]* and *TALEN-R[S1]* DNA including the seventeen 102-bp tandem DNA repeats. The position (arrowheads) and size (in kb) of the 2.1-kb amplicon diagnostic for intact *TALEN* transgenes are indicated at the right. Marker, Gene Ruler DNA Ladder Mix. H$_2$0, PCR performed with nuclease-free water instead of vector DNA template. **(D)** Autoradiogram of *Mun*I-treated extrachromosomal DNA isolated from HeLa cells transduced with Ad.ΔE1.TALEN-L[S1].F[50] or with Ad.ΔE1.TALEN-R[S1].F[50] at MOIs of 10, 5 and 1 HTU/cell. The DNA, extracted at three days post-transduction, was subjected to Southern blot analysis using a 397-bp probe hybridizing to a sequence encoding the NH2 terminus of the TALEN-L[S1] and TALEN-R[S1] proteins. ▶

A

Ad.ΔE1.TALEN-L^{S1}.F^{50}
(MOI = 6 HTU/cell)

Ad.ΔE1.TALEN-R^{S1}.F^{50}
(MOI = 6 HTU/cell)

Ad.ΔE1.TALEN-L^{S1}.F^{50}
+
Ad.ΔE1.TALEN-R^{S1}.F^{50}
(MOI = 6+6 HTU/cell)

phase

RFP

HeLa

Ad.ΔE1.TALEN-L^{S1}.F^{50}
(MOI = 12 HTU/cell)

Ad.ΔE1.TALEN-R^{S1}.F^{50}
(MOI = 12 HTU/cell)

Ad.ΔE1.TALEN-L^{S1}.F^{50}
+
Ad.ΔE1.TALEN-R^{S1}.F^{50}
(MOI = 12+12 HTU/cell)

RFP

Myoblasts

hMSCs

B

mock

Ad.ΔE1.TALEN-L^{S1}.F^{50}
(MOI = 3 HTU/cell)

Ad.ΔE1.TALEN-R^{S1}.F^{50}
(MOI = 3 HTU/cell)

Hoechst

HA::TALEN

HeLa

C

HTU/cell

4   2   1   0.5   0.25   0

TALEN-L^{S1} →

TALEN-R^{S1} →

HA::TALEN

Finally, to investigate the functionality of TALENs introduced into human cells through adenoviral vector-mediated gene transfer, we deployed the Surveyor mutation detection assay (Figure 4A). PCR amplification of *AAVS1* target DNA from mock-transduced cells or from cells transduced with Ad.ΔE1.TALEN-L^{S1}.F^{50} or with Ad.ΔE1.TALEN-R^{S1}.F^{50} yielded amplicons that, upon denaturation and reannealing, were not susceptible to site-specific cleavage by the mismatch-sensitive SNS enzyme (Figure 4B, solid arrowhead in the upper panel). This data validated the read-out system by demonstrating the absence of indels at the *AAVS1* locus under these experimental control conditions. However, co-transduction of HeLa cells and DMD myoblasts with Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} resulted in efficient cleavage of the endogenous chromosomal target sequence, as evidenced by the detection of large amounts of SNS-digested products (Figure 4B, open arrowheads in the lower panels). This conclusion was substantiated through vector dose-response experiments on HeLa cells, human myoblasts and hMSCs (Figure 4C). Additional transduction experiments on hMSCs that made use of the second-generation adenoviral vectors Ad.ΔE1E2A.TALEN-L^{S1}.F^{50} and Ad.ΔE1E2A.TALEN-R^{S1}.F^{50} produced results similar to those shown in Figure 4C (Supplementary Figure S4). Finally, we sought to confirm the involvement of the NHEJ pathway in the repair of the TALEN-induced DSBs. To this end, we cloned the *AAVS1*-specific PCR products resulting from genomic DNA of hMSCs exposed to Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} at an MOI of 12 HTU/cell each. DNA sequencing of twenty randomly selected clones revealed that 12 of these harbored indels characteristic of NHEJ-mediated repair of DSBs (Figure 4D).

We conclude that delivery of TALENs by first- and second-generation adenoviral vectors is a highly effective approach to generate in a cell type-independent manner site-specific DSBs at an endogenous human target locus.

Figure 3. Adenoviral vector-mediated delivery of *TALEN* genes into transformed and nontransformed human cells. **(A)** Gene transfer efficiency. Upper panel, phase-contrast and live-cell RFP direct fluorescence microscopy on HeLa cells transduced with Ad.ΔE1.TALEN-L^{S1}.F^{50} or with Ad.ΔE1.TALEN-R^{S1}.F^{50} at an MOI of 6 HTU/cell or with a mixture of the two adenoviral vectors at an MOI of 6 HTU/cell each. Lower panel, RFP direct fluorescence microscopy on immortalized DMD myoblasts and on primary hMSCs transduced with Ad.ΔE1.TALEN-L^{S1}.F^{50} or with Ad.ΔE1.TALEN-R^{S1}.F^{50} at an MOI of 12 HTU/cell or co-transduced with a 1:1 mixture of Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} at a total MOI of 24 HTU/cell. The fluorescence microscopy images were acquired at 48 hours post-infection. **(B)** HA tag-specific immunofluorescence microscopy on HeLa cells mock-transduced or transduced with Ad.ΔE1.TALEN-L^{S1}.F^{50} or with Ad.ΔE1.TALEN-R^{S1}.F^{50} at an MOI of 3 HTU/cell. The direct fluorescence microscopy and the immunofluorescence microscopy were carried out at 48 and 72 hours post-infection, respectively. **(C)** HA tag-directed Western blot analysis of protein lysates derived from HeLa cell cultures incubated for 72 hours with Ad.ΔE1.TALEN-L^{S1}.F^{50} or with Ad.ΔE1.TALEN-R^{S1}.F^{50}. The MOIs deployed are indicated.
◄

Figure 4. Targeted disruption of the chromosomal *AAVS1* locus. **(A)** Schematic representation of the Surveyor mutation detection assay used to probe TALEN activity in human cells. Assembly of the TALEN pair at the bipartite target sequence results in local DSB formation. The errorprone NHEJ DNA repair pathway links the broken chromosomal ends introducing in the process small insertions and deletions (indels) at the junction (grey boxes). PCR on chromosomal DNA with primers framing the target site yields a mixture of amplicons representing the heterogeneous population of mutated sequences and non-modified alleles. Denaturation and re-annealing of this mixture results in DNA species with internal base pairs that are mismatched. The Surveyor nuclease S (SNS) enzyme, by preferentially recognizing and cleaving at DNA sequences with such mismatched base pairs, yields low-molecular-weight DNA fragments (open arrowheads) that can be resolved from undigested DNA (solid arrowheads) through conventional agarose gel electrophoreses. Parallel samples not treated with SNS (-) served as a negative control. **(B)** TALEN activity at *AAVS1*. Upper panel, Surveyor assay on chromosomal DNA isolated from HeLa cells not exposed (mock) or exposed to Ad.ΔE1.TALEN-L^{S1}.F^{50} or to Ad.ΔE1.TALEN-R^{S1}.F^{50} at a high and a low dose (i.e. 6 and 2 HTU/cell, respectively). Lower panels, Surveyor assay on chromosomal DNA extracted from HeLa cells and DMD myoblasts (left and right panels, respectively). HeLa cells were mock-transduced (-) or were co-transduced with 6 HTU/cell of Ad.ΔE1.TALEN-L^{S1}.F^{50} and 6 HTU/cell of Ad.ΔE1.TALEN-R^{S1}.F^{50} (+). Human myoblasts were mock-transduced (-) or were co-transduced with 12 HTU/cell of Ad.ΔE1.TALEN-L^{S1}.F^{50} and 12 HTU/cell of Ad.ΔE1.TALEN-R^{S1}.F^{50} (+). Solid and open arrowheads indicate the positions of undigested and SNS-digested DNA molecules, respectively. White numerals correspond to the percentage of target site cleavage. **(C)** Dose-response of TALEN activity. Surveyor assay on cellular DNA from HeLa cells, DMD myoblasts and hMSCs untreated or treated with different doses of 1:1 mixtures of Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50}. In these dose-response co-transduction experiments the total amounts of vector applied on HeLa cells were 12, 4, 1.3, 0.4 and 0.15 HTU/cell, whereas on DMD myoblasts and hMSCs HTU/cell were 24, 8, 2.7, 1.0 and 0.3. Solid and open arrowheads indicate the positions of undigested and SNS-digested DNA molecules, respectively. White numerals correspond to the percentage of target site cleavage. **(D)** Sequence analysis. DNA sequencing of cloned *AAVS1*-specific amplicons resulting from PCR amplifications on genomic DNA isolated from hMSCs co-transduced with Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} at 12 HTU/cell each. The presence of indels in 12 out of 20 clones, "footprint" for NHEJ processing of DSBs, demonstrates at the nucleotide level the TALEN-mediated cleavage of the chromosomal *AAVS1* target site. ▶
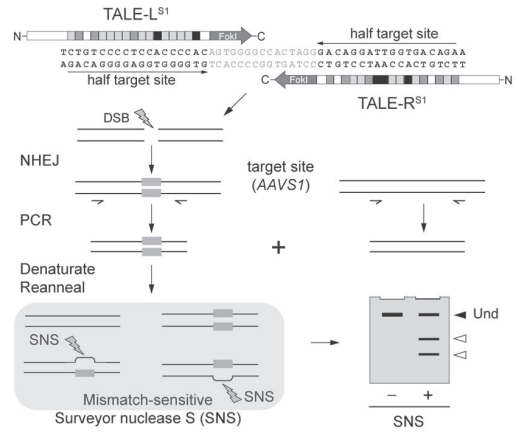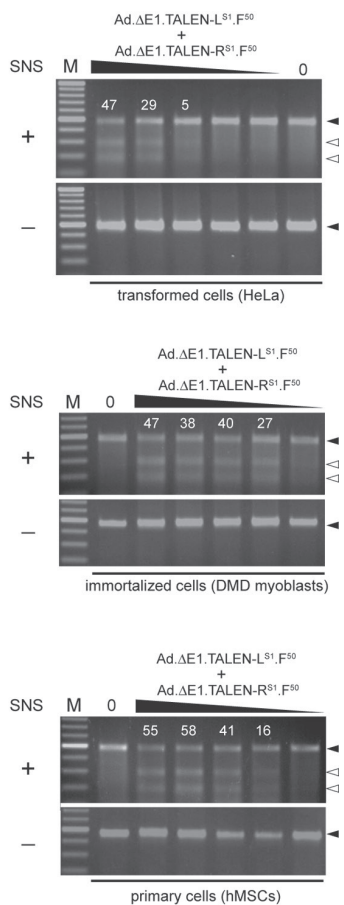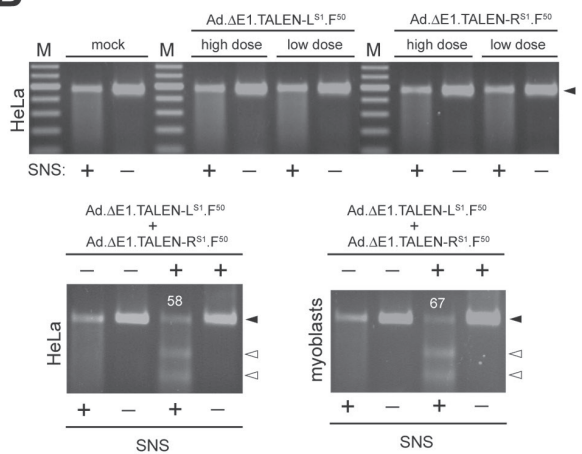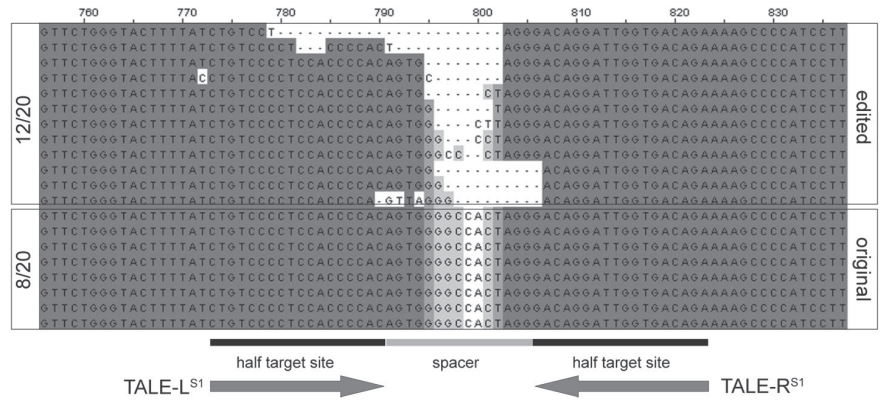
DISCUSSION

The approaches aiming at editing or modulating the expression of specific endogenous genes and targeting chromosomal insertion of exogenous DNA, have recently been enriched with the introduction of sequence-specific designer TALEs. To broaden the use of TALE-based proteins and to most optimally screen and evaluate their performance in many settings, it will be crucial to expand the range of cellular delivery tools for these molecules. Despite their promise and favorable set of attributes when compared to other custom-made sequence-specific proteins, TALEs are relatively large proteins whose DNA binding specificity is governed by an extensive tract of repetitive sequences. However, DNA repeats are known to be particularly unstable motifs. Indeed, examples of both physiological and function-impairing processes associated with instability of tandem repeats are copious and cover virtually all biological systems from viruses and prokaryotes to eukaryotes (44-46). Also in the context of gene delivery vehicles, such as those based on adenoviral and lentiviral vectors, examples exist pointing to certain repetitive sequence arrangements as the culprits behind DNA instability (47-50). Bearing this in mind, in the present study we sought to investigate the suitability of adenoviral and lentiviral vectors, two of the most commonly used viral vector systems, to serve as *TALEN* gene delivery platforms.

We report that lentiviral vector genomes bearing *TALEN* sequences are prone to rearrangements in target cells. In this regard, the structural analyses of these genomes in HeLa cell populations and in individual HeLa cell clones by PCR and Southern blot analyses, respectively, suggest that most of the rearrangements occurred through recombination events involving the *TALE* repeat array ultimately leading to deletions with various sizes. Conversely, we found that *TALEN* sequences are stably maintained in first- and second-generation adenoviral vectors following their serial propagation in producer cells. Indeed, the *TALEN* transgenes in the adenoviral vector genomes presented neither large rearrangements nor small-scale mutations as revealed by restriction fragment length analysis and DNA sequencing, respectively. Importantly, transduction experiments with the TALEN-encoding adenoviral vectors led to DSB formation at the intended target chromosomal locus in transformed and non-transformed human cells at similarly high frequencies.

Clearly, viruses from the *Adenoviridae* and *Retroviridae* families display diametrically different structural features, life cycles and replication strategies (15,51,52). The latter aspect in particular is likely to contribute to the disparate stability of *TALEN* ORFs in adenoviral versus lentiviral vectors. Adenoviruses replicate via a DNA strand-displacement mechanism using a virus-encoded high-fidelity DNA polymerase and a single-stranded DNA-binding protein that, by cooperatively coating the displaced single-stranded replicative intermediates, presumably inhibits spurious intramolecular hybridization (52,53). Retroviruses, on the other hand, encode error-prone reverse transcriptases (RTs), which convert their double-

copy RNA genomes into linear double-stranded DNA templates that serve as substrates for proviral chromosomal insertion (51). The RTs of simple and complex retroviruses also display a RNase H activity responsible for the digestion of the RNA component of hybrid RNA:DNA duplex intermediates. Moreover, they share a low template affinity and processivity. These features seem to have evolved to allow RT to undergo at least two obligate template switching events (i.e. minus-strand and plus-strand DNA transfer) in order to complete full-length viral DNA synthesis and, as "byproduct", generate genotypic diversity through recombination. It is known, however, that the presence of direct repeats in retroviral genomes increases the likelihood for additional inter- and intra-molecular template switching events, often resulting in the deletion of one of the repeats plus intervening sequences (47,48,54,55). In fact, the insertion of direct repeats in retroviral genomes provides for an elegant model system to study the RT template switching process in vivo or to induce precise rearrangements of *cis*- and *trans*-acting elements in retroviral vectors (54,56). Of note, the frequency of template switching increases with the size of and the distance between the direct repeat units. The latter finding suggests that spacing favors template "looping" which, in turn, aids in the intra-strand alignment of direct repeats, RT template switching and ensuing intra-molecular deletions (55). The models for these direct repeat-induced deletions (48,54,55) postulate that during reverse transcription, RT and the nascent DNA strand dissociates from the RNA template due to the aforementioned RNaseH activity and intrinsic low template affinity. Segments and, eventually, all nascent DNA strand hybridizes to the RNA template at the distal "acceptor" repeat serving as a novel "ectopic" initiation point for RNA-dependent DNA synthesis. The net result of these processes is often homogenously sized deletions of the retroviral genome. In contrast to simple direct repeat pairs, *TALE* sequences present a tract of numerous and contiguous repeats. Thus, each of these arrayed repeat units provide for multiple potential positions at which RT template switching may occur. This corollary is consistent with the heterogeneous and sub-genomic sized population of TALENcontaining lentiviral vector proviruses that were detected in stably transduced clones.

Notwithstanding these findings, one can envision strategies that might minimize or overcome the instability of *TALEN* sequences in lentiviral vectors. For instance, by exploiting the degeneracy of the genetic code, the individual TALE motifs within the array could be reengineered by the introduction of judiciously chosen silent substitutions to increase sequence divergence among repeats. Regarding this issue, it has been shown via a lentiviral vector direct repeat deletion assay that deletion frequencies can decrease steeply with incremental repeat sequence divergence (57).

In conclusion, *TALEN* sequences are prone to rearrangements following their introduction into human cells by lentiviral vectors. As a result, effective lentiviral vectormediated *TALEN* gene transfer will be dependent upon new optimization strategies. Conversely, we demonstrate that first- and second-generation adenoviral vectors are amenable to the

transfer of full-length *TALEN* genes into human cells. Importantly, the resulting TALEN protein yields suffice to induce high-level DSB formation at an endogenous chromosomal target locus.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online: Supplementary Figures 1-4.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA*, **93**, 1156-1160.

2. Klug,A. (2010) The discovery of Zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem*., **79**, 213-231.

3. Redondo,P., Prieto,J., Muñoz,I.G., Alibés,A., Stricher,F., Serrano,L., Cabaniols,J.P., Daboussi,F., Arnould,S., Perez,C., Duchateau,P., Pâques,F., Blanco,F.J. and Montoya,G. (2008) Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature*, **456**, 107-111.

4. Silva.G., Poirot,L., Galetto,R., Smith,J., Montoya,G., Duchateau,P. and Pâques,F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther*., **11**, 11-27.

5. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TALtype III effectors. *Science*, **326**, 1509-1512.

6. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 150.

7. Boch,J. and Bonas,U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol*., **48**, 419-436.

8. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J., Dulay,G.P., Hua,K.L., Ankoudinova,I., Cost,G.J., Urnov,F.D., Zhang,H.S., Holmes,M.C., Zhang,L., Gregory,P.D. and Rebar,E.J. (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol*., **29**, 143-148.

9. Mussolino,C., Morbitzer,R., Lütge,F., Dannemann,N., Lahaye,T. and Cathomen,T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*., **39**, 9283-9293.

10. Mussolino,C. and Cathomen,T. (2012) TALE nucleases: tailored genome engineering made easy. *Curr. Opin. in Biotech*., **23**, 1-7.

11. Cathomen,T. and Joung,K.J. (2008) Zinc-finger nucleases: the next generation emerges. *Mol. Ther*., **16**, 1200-1207.

12. Hockemeyer,D., Wang,H., Kiani,S., Lai,C.S., Gao,Q., Cassady,J.P., Cost,G.J., Zhang,L., Santiago,Y., Miller,J.C., Zeitler,B., Cherone,J.M., Meng,X., Hinkley,S.J., Rebar,E.J., Gregory,P.D., Urnov,F.D. and Jaenisch,R. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol*., **29**, 731-734.

13. Tong,C., Huang,G., Ashton,C., Wu,H., Yan,H. and Ying,Q.L. (2012) Rapid and costeffective gene targeting in rat embryonic stem cells by TALENs. *J. Genet. Genomics*, **39**, 275-280.

14. Sun,N., Liang,J., Abil,Z. and Zhao,H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. Biosyst*., **8**, 1255-1263.

15. Gonçalves,M.A.F.V. and de Vries,A.A.F. (2006) Adenovirus: from foe to friend. *Rev. Med. Virol*., **16**, 167-186.

16. Dropulić,B. (2011) Lentiviral vectors: their molecular design, safety, and use in laboratory and preclinical research. *Hum. Gene Ther*., **22**, 649-657.

17. Yamashita,M. and Emerman,M. (2006) Retroviral infection of non-dividing cells: old and new perspectives. *Virology*, **344**, 88-93.

18. Henaff,D., Salinas,S. and Kremer,E.J. (2011) An adenovirus traffic update: from receptor engagement to the nuclear pore. *Future Microbiol*., **6**, 179-192.

19. Fallaux,F.J., Bout,A., van der Velde,I., van den Wollenberg,D.J., Hehir,K.M., Keegan,J., Auger,C., Cramer,S.J., van Ormondt,H., van der Eb,A.J., Valerio,D. and Hoeben,R.C. (1998) New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Hum. Gene Ther.*, **9**, 1909-1917.

20. Havenga,M.J., Holterman,L., Melis,I., Smits,S., Kaspers,J., Heemskerk,E., van der Vlugt,R., Koldijk,M., Schouten,G.J., Hateboer,G., Brouwer,K., Vogels,R. and Goudsmit,J. (2008) Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: high yield and preserved bioactivity. *Biotechnol. Bioeng.*, **100**, 273-283.

21. Cudré-Mauroux,C., Occhiodoro,T., König,S., Salmon,P., Bernheim,L. and Trono,D. (2003) Lentivector-mediated transfer of Bmi-1 and telomerase in muscle satellite cells yields a duchenne myoblast cell line with long-term genotypic and phenotypic stability. *Hum. Gene Ther.*, **14**, 1525-1533.

22. Gonçalves,M.A.F.V., Janssen,J.M., Nguyen,Q.G., Athanasopoulos,T., Hauschka,S.D., Dickson,G. and de Vries,A.A.F. (2011) Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. *Mol. Ther.*, **19**, 1331-1341.

23. Janssen,J.M., Liu,J., Skokan,J., Gonçalves,M.A.F.G. and de Vries,A.A.F. (2012) Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. *Submitted*.

24. Seppen,J., Rijnberg,M., Cooreman,M.P. and Oude Elferink,R.P. (2002) Lentiviral vectors for efficient transduction of isolated primary quiescent hepatocytes. *J. Hepatol.*, **36**, 459-465.

25. Pelascini,P.L., Janssen,J.M. and Gonçalves,M.A.F.V. (2012) Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. Hum. Gene Ther. *In press*.

26. Gonçalves,M.A.F.V., Swildens,J., Holkers,M., Narain,A., van Nierop,G.P., van de Watering,M.J., Knaän-Shanzer,S. and de Vries,A.A.F. (2008) Genetic complementation of human muscle cells via directed stem cell fusion. *Mol. Ther.*, **16**, 741-748.

27. Gonçalves,M.A.F.G., Holkers,M., van Nierop,G.P., Wieringa,R., Pau,M.G. and de Vries,A.A.F. (2008) Targeted chromosomal insertion of large DNA into the human genome by a fiber-modified high-capacity adenovirus-based vector system. *PLoS One*, **29**, e3084.

28. van Nierop,G.P., de Vries,A.A.F., Holkers,M., Vrijsen,K.R. and Gonçalves,M.A.F.V. (2009) Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Res.*, **37**, 5725-5736.

29. Gonçalves,M.A.F.V., van der Velde,I., Knaän-Shanzer,S., Valerio,D. and de Vries,A.A.F. (2004) Stable transduction of large DNA by high-capacity adeno-associated virus/adenovirus hybrid vectors. *Virology*, **321**, 287-296.

30. Szulc,J., Wiznerowicz,M., Sauvain,M.O., Trono,D. and Aebischer,P. (2006) A versatile tool for conditional gene expression and knockdown. *Nat. Methods*, **3**, 109-116.

31. Philpott,N.J. and Thrasher,A.J. (2007) Use of nonintegrating lentiviral vectors for gene therapy. *Hum. Gene Ther.*, **18**, 483-489.

32. Wanisch,K. and Yáñez-Muñoz,R.J. (2009) Integration-deficient lentiviral vectors: a slow coming of age. *Mol. Ther.*, **17**, 1316-1332.

33. Staunstrup,N.H. and Mikkelsen,J.G. (2011) Integrase-defective lentiviral vectors-a stage for nonviral integration machineries. *Curr. Gene Ther.*, **11**, 350-362.

34. Cornu,T.I. and Cathomen,T. (2007) Targeted genome modifications using integrasedeficient lentiviral vectors. *Mol. Ther.*, **15**, 2107-2113.

35. Lombardo,A., Genovese,P., Beausejour,C.M., Colleoni,S., Lee,Y.L., Kim,K.A., Ando,D., Urnov,F.D., Galli,C., Gregory,P.D., Holmes,M.C. and Naldini,L. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.*, **25**, 1298-1306.

36. Izmiryan,A., Basmaciogullari,S., Henry,A., Paques,F. and Danos,O. (2011) Efficient gene targeting mediated by a lentiviral vector-associated meganuclease. *Nucleic Acids Res*., **39**, 7610-7619.

37. Shayakhmetov,D.M., Papayannopoulou,T., Stamatoyannopoulos,G. and Lieber,A. (2000) Efficient gene transfer into human CD34(+) cells by a retargeted adenovirus vector. *J. Virol*., **74**, 2567-2583.

38. Knaän-Shanzer,S., van der Velde,I., Havenga,M.J., Lemckert,A.A., de Vries,A.A.F. and Valerio,D. (2001) Highly efficient targeted transduction of undifferentiated human hematopoietic cells by adenoviral vectors displaying fiber knobs of subgroup B. *Hum. Gene Ther*., **12**, 1989-2005. Erratum in: *Hum. Gene Ther*. 2003 Aug 10;14.

39. Knaän-Shanzer,S., van de Watering,M.J., van der Velde,I., Gonçalves,M.A.F.V., Valerio,D. and de Vries,A.A.F. (2005) Endowing human adenovirus serotype 5 vectors with fiber domains of species B greatly enhances gene transfer into human mesenchymal stem cells. *Stem Cells*, **23**, 1598-1607.

40. Tuve,S., Wang,H., Ware,C., Liu,Y., Gaggar,A., Bernt,K., Shayakhmetov,D., Li,Z., Strauss,R., Stone,D. and Lieber,A. (2006) A new group B adenovirus receptor is expressed at high levels on human stem and tumor cells. *J. Virol*., **80**, 12109-12120.

41. Hoffmann,D., Meyer,B. and Wildner,O. (2007) Improved glioblastoma treatment with Ad5/35 fiber chimeric conditionally replicating adenoviruses. *J. Gene Med*., **9**, 764-778.

42. Gonçalves,M.A.F.V., Holkers,M., Cudré-Mauroux,C., van Nierop,G.P., Knaän-Shanzer,S., van der Velde,I., Valerio,D. and de Vries,A.A.F. (2006) Transduction of myogenic cells by retargeted dual high-capacity hybrid viral vectors: robust dystrophin synthesis in duchenne muscular dystrophy muscle cells. *Mol. Ther*., **13**, 976-986.

43. Gonçalves,M.A.F.V., de Vries,A.A.F., Holkers,M., van de Watering,M.J., van der Velde,I., van Nierop,G.P., Valerio,D. and Knaän-Shanzer,S. (2006) Human mesenchymal stem cells ectopically expressing full-length dystrophin can complement Duchenne muscular dystrophy myotubes by cell fusion. *Hum. Mol. Genet*., **15**, 213-221.

44. Pearson,C.E., Nichol Edamura,K. and Cleary,J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet*., **6**, 729-742.

45. Zhao,J., Bacolla,A., Wang,G. and Vasquez,K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci*., **67**, 43-62.

46. Holkers,M., de Vries,A.A.F. and Gonçalves,M.A.F.V. (2012) Nonspaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells. *Nucleic Acids Res.*, **40**, 1984-1999.

47. Rhode,B.W., Emerman,M. and Temin,H.M. (1987) Instability of large direct repeats in retrovirus vectors. *J. Virol*., **61**, 925-927.

48. Pathak,V.K. and Temin,H.M. (1990) Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc. Natl. Acad. Sci. USA*, **87**, 6024-6028.

49. Steinwaerder,D.S., Carlson,C.A. and Lieber,A. (1999) Generation of adenovirus vectors devoid of all viral genes by recombination between inverted repeats. *J. Virol*., **73**, 9303-9313.

50. Belousova,N., Harris,R., Zinn,K., Rhodes-Selser,M.A., Kotov,A., Kotova,O., Wang,M., Aurigemma,R., Zhu,Z.B., Curiel,D.T. and Alvarez,R.D. (2006) Circumventing recombination events encountered with production of a clinical-grade adenoviral vector with a double-expression cassette. *Mol. Pharmacol*., **70**, 1488-1493.

51. Coffin,J.M., Hughes,S.H. and Varmus,H.E. (1997) Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

52. de Jong,R.N., van der Vliet,P.C. and Brenkman,A.B. (2003) Adenovirus DNA replication: protein priming, jumping back and the role of the DNA binding protein DBP. *Curr Top Microbiol Immunol.*, **272**, 187-211.

3

53. Zijderveld,D.C., Stuiver,M.H. and van der Vliet,P.C. (1993) The adenovirus DNA binding protein enhances intermolecular DNA renaturation but inhibits intramolecular DNA renaturation. *Nucleic Acids Res*., **21**, 2591-2598.

54. Julias,J.G., Hash,D. and Pathak,V.K. (1995) E- vectors: development of novel selfinactivating and self-activating retroviral vectors for safer gene therapy. *J. Virol*., **69**, 6839-6846.

55. Delviks,K.A. and Pathak,V.K. (1999) Effect of distance between homologous sequences and 3' homology on the frequency of retroviral reverse transcriptase template switching. *J. Virol*., **73**, 7923-7932.

56. Delviks,K.A., Hu,W.S. and Pathak,V.K. (1997) Psi- vectors: murine leukemia virusbased self-inactivating and self-activating retroviral vectors. *J. Virol*., **71**, 6218-6224.

57. An,W. and Telesnitsky,A. (2002) Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J. Virol*., **76**, 7897-7902.

3

Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells

4

Maarten Holkers[1], Toni Cathomen[2,3], Manuel A.F.V. Gonçalves[1]

[1]Department of Molecular Cell Biology, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands; [2]Institute for Cell and Gene Therapy and [3]Center for Chronic Immunodeficiency, University Medical Center Freiburg, 79106 Freiburg, Germany.

ABSTRACT

Transcription activator-like effector nucleases (TALENs) are designed to cut the genomic DNA at specific chromosomal positions. The resulting DNA double strand break activates cellular repair pathways that can be harnessed for targeted genome modifications. TALENs thus constitute a powerful tool to interrogate the function of DNA sequences within complex genomes. Moreover, their high DNA cleavage activity combined with a low cytotoxicity make them excellent candidates for applications in human gene therapy. Full exploitation of these large and repeat-bearing nucleases in human cell types will benefit largely from using the adenoviral vector (AdV) technology. The genetic stability and the episomal nature of AdV genomes in conjunction with the availability of a large number of AdV serotypes able to transduce various human cell types make it possible to achieve high-level and transient expression of TALENs in numerous target cells, regardless of their mitotic state. Here, we describe a set of protocols detailing the rescue, propagation and purification of TALEN-encoding AdVs. Moreover, we describe procedures for the characterization and quantification of recombinant viral DNA present in the resulting AdV preparations. The protocols are preceded by information about their underlying principles and applied in the context of second-generation capsid-modified AdVs expressing TALENs targeted to the *AAVS1* "safe harbor" locus on human chromosome 19.

INTRODUCTION

Targeted genome engineering is based on the creation of double-stranded DNA breaks (DSBs) at predefined chromosomal positions, which activate DNA repair pathways such as non-homologous end-joining (NHEJ) or the homologous recombination (HR) machinery (1). The potential sphere of action of genome engineering is broad and includes the development of new molecular medicine modalities based on the addition of therapeutic genes at so-called "safe harbor" loci or the correction of endogenous defective genes. Ultimately, such "genome surgery" should constitute a departure from conventional gene therapy, in which exogenous DNA is inserted at unpredictable chromosomal positions (2).

Generically, genome editing strategies depend on the introduction of designer nucleases into target cells for inducing the site-specific DSBs (3). Designer nuclease technologies are evolving at a rapid pace and include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and, more recently, RNA-guided nucleases (RGNs) based on the prokaryotic CRISPR/Cas9 adaptive immune systems (4). While each of these platforms have their pros and cons, assessment of the specificities of ZFNs and RGNs revealed a considerable degree of off-target DNA cleavage (5-9). On the other hand, currently available data suggest that TALENs seem to be better tolerated in mammalian cells than ZFNs (10,11), although it is not clear yet whether "better tolerability" correlates with higher target site specificity, a key parameter in future therapeutic applications of genome editing protocols in patient-derived cells.

TALENs are easy to engineer heterodimeric designer nucleases that combine high cleavage activity with low cytotoxicity (12). The modularly structured DNA binding domain of transcription activator-like effectors (TALEs) was originally discovered in pathogenic bacteria of the genus *Xanthomonas* (13,14). Each repeat module within the DNA recognition domain consists of a conserved stretch of about 34 residues. The interaction with a single nucleotide of the DNA target site is mediated via two variable residues in positions 12 and 13. Repeat modules with different specificities can be fused into tailored arrays to direct the non-specific *Fok*I endonuclease domain to a predetermined target site (15,16).

Adenoviral vectors (AdVs) based on human adenoviruses (HAdVs) are widespread and versatile vehicles for introducing recombinant genes into human cells (17). Among these, those encoding immunomodulatory proteins or antigens are often selected as therapeutic payloads for anticancer and vaccination purposes (18). The broad use and adaptability of AdVs stem from several key features of their parental viruses as well as the extensive knowledge gathered on their biology since they were discovered in 1953 (19).

Presently, over 50 human adenovirus serotypes have been identified and classified in various sub-groups or species (A to G). They are mildly pathogenic viruses consisting of a linear protein-capped double-stranded DNA genome packaged in a non-enveloped

icosahedral capsid with fiber proteins protruding from each of its twelve vertices (17,18). The structural proteins are involved in host cell attachment by binding to specific cell surface receptors and can differ substantially among the various serotypes (20). As a result, the cell tropism of *Adenoviridae* family members can vary profoundly, owing in large part to their particular fiber-receptor interactions. For instance, under regular cell culture conditions, the prototypic species C serotypes 2 and 5 utilize the Coxsackievirus B and Adenovirus receptor (CAR) as primary cell surface attachment moiety (21), whereas CD46 is engaged by, amongst others, the species B serotypes 35 and 50 (22). An important feature of human adenoviruses is the similarity with which they infect diving and non-dividing cells. Indeed, the entry of viral genomes into host cell nuclei is independent of nuclear envelope integrity owing to the active process by which endosome-remodeled virions dock at the nuclear pore complex, disassemble and release their genomes into the nucleoplasm after having been transported in complex with dynein along the cytoplasmic microtubule network (23). Once in the nucleus, the orchestrated activation of the "early" (*E*) and "late" (*L*) viral gene expression programs yields large amounts of newly assembled mature particles. At the completion of the lytic viral life cycle, 10^3 to 10^4 virions are typically found within each host cell nucleus.

The most common AdVs are based on the CAR-interacting prototypic HAdV serotype 5 and are deleted in *E* regions whose products participate in viral DNA replication and/or activate the viral gene expression program (17). In addition to disrupting the regular replication cycle in target cells, these deletions create space for the insertion of exogenous DNA. As a result, for their production, these vectors require packaging cell lines, expressing *in trans* the viral genes, which have been deleted in the AdV backbone. These types of AdVs are called helper-independent as they can be amplified autonomously in such specialized packaging cell lines. The simplest helper-independent AdV type lacks the viral early region 1 (*E1*) and are thus called *E1*-deleted or first-generation AdVs. Notwithstanding the attributes of first-generation AdVs and their suitability for many different applications, it was shown that at high multiplicities of infection (MOI), the removal of *E1* does not fully block expression of viral open reading frames (ORFs), presumably due to partially complementing *E1A*-like cellular functions (24). Albeit at low-levels, the ensuing *de novo* viral protein synthesis can hamper the optimal deployment of these AdVs in certain experimental settings, such as those involving the transduction of therapeutically relevant cell types. This has prompted the development of second-generation AdVs combining deletions in *E1* with those in other *E* region(s), namely, early region 2A (*E2A*) and/or early region 4 (*E4*) (17). Of note, because the functions encoded in early region 3 (*E3*) are dispensable for vector propagation in cells cultured *in vitro*, first- and second-generation AdVs are often also deleted in *E3*. In addition to curtailing the phenomenon of "leaky" viral gene expression in target cells, propagation of second-generation AdVs is less likely to yield contaminating replication-competent AdVs (RCAs) resulting from recombination events between viral sequences shared by vector

and commonly used producer-cell genomes (25). Clearly, the issue of RCA contamination is particularly relevant whenever AdV preparations are produced for therapeutic gene transfer protocols (26). For this reason, AdV packaging cell lines, such as PER.C6 and N52.E6, were designed without DNA sequence overlap between the recombinant portions of their genomes and AdV DNA (27,28).

The available in-depth knowledge about adenovirus capsid structure and gene regulation allows constructing AdVs with genomic modifications not only in *E* but also in *L* ORFs. For instance, genetic manipulation of the fiber-encoding *L* transcription unit 5 (*L5*) permits retargeting AdV particles to cells with low or no CAR on their surface. These transductional targeting strategies can be based on engineering capsids with heterologous ligands (29) or on tapping the natural diversity of HAdVs to create chimeric fibers displaying receptor-interacting motifs derived from a serotype with a selected tropism of choice (29,30). This "fiber swapping" approach is based on exchanging the apical fiber motifs of serotype 5 for those of another serotype, such as serotype 35 or serotype 50, and has proven to be very valuable for the transduction of therapeutically relevant CAR-negative cells. These cells include normal and malignant hematopoietic cells as well as myocytes and mesenchymal stromal cells (20,29,30).

We postulate that AdV technologies, besides their utility in established biomedical research areas such as oncology and vaccinology, will start playing an increasing role in the emerging field of genome engineering as well. In fact, AdV-mediated introduction of ZFNs into CD4[+] T cells to knockout the HIV co-receptor gene *CCR5*, represents the first testing of a designer nuclease-based therapeutic strategy in humans (31). Related to this, our laboratories together with collaborators have recently demonstrated that AdVs can introduce functional *TALEN* ORFs into human cells (32). Importantly, the frequencies of targeted mutagenesis achieved by AdV-mediated TALEN expression were similar in transformed and non-transformed cells (32). In contrast, the same *TALEN* ORFs delivered in the context of conventional and integration-defective lentiviral vectors suffered from extensive function-impairing rearrangements within their repeat arrays, presumably due to frequent template switching of the reverse transcriptase (32). Yang and colleagues have subsequently confirmed these results by demonstrating that lentiviral vector-mediated delivery of intact *TALEN* ORFs requires substantial sequence optimization to reduce the length and frequency of repetitive sequences (33). These optimized TALENs were dubbed re-coded TALENs. However, because there are precedents for the importance in achieving high yields of sequence-specific nucleases in target cells for robust site-specific DSB formation (34,35), AdVs might ultimately be preferable over silencing-prone IDLVs (36) for transient TALEN expression.

In the current article, we provide a set of protocols detailing the production and purification of AdVs encoding TALENs. In addition, we describe in a comprehensive manner

procedures for the characterization and quantification of AdV genomes. These protocols are applied in the context of producing and validating a pair of second-generation fiber-modified AdVs containing *TALEN* ORFs under the control of the house-keeping *PGK1* gene promoter (PGK). The resulting TALEN proteins are addressed to target sequences within the so-called *AAVS1* "safe harbor" locus, located in the long arm of the human chromosome 19 at position 19q13.42-qter. Of note, the herein described methodologies can be directly applied to the generation and characterization of AdV particles harboring other transgenes of interest, including those encoding other types of sequence-tailored nucleases or genome-modifying enzymes.

CONSTRUCTION OF ADV MOLECULAR CLONES

Materials
1. Electro-competent *E. coli* strain BJ5183 (Addgene bacterial strain 16398) containing the pAdEasy-1 "backbone" plasmid or derivatives (here BJ5183[pAdEasy-2.50])
2. *E. coli* strain DH5α (Life Technologies)
3. pAdShuttle expression plasmids (e.g. pAdSh.PGK.SV40)
4. TALEN-encoding plasmids (e.g. 1383.pVAX.AAVS1.TALEN.L-94 and 1384.pVAX. AAVS1.TALEN.R-95)
5. 10 mg/ml transfer RNA (tRNA) carrier (Sigma-Aldrich)
6. Luria-Bertani broth (LB; Life Technologies) medium and agar plates
7. 500 µg/ml kanamycin sulfate (Sigma-Aldrich)
8. PmeI (New England BioLabs) or MssI (Thermo Scientific)
9. Absolute ethanol (J.T. Baker)
10. Table-top Eppendorf centrifuge
11. Electroporator (here Gene Pulser electroporator [Bio-Rad])
12. Agarose gel electrophoresis reagents and apparatus
13. Plasmid DNA purification system e.g. JETSTAR 2.0 Plasmid Maxiprep Kit (Genomed)
14. Spectrophotometer (here NanoDrop ND-1000 [Thermo Scientific])

Methods
The methods for constructing helper-independent AdV genomes are numerous. Among the first were those based on assembling full-length AdV DNA in producer human cells by HR between co-transfected "shuttle" and "backbone" constructs. Alternative methods based on AdV genome assembly in yeast or bacteria cells were subsequently developed due to the inefficiency and unpredictability of DNA recombination in human cells (17). Among these procedures, the AdEasy system is a particularly simple and robust one owing to its reduced

dependency on enzymatic manipulations and reproducibility in recovering plasmids containing whole AdV DNA (37). From the onset these reagents correspond to individual molecular clones that can be directly applied for rescuing and propagating AdV particles in producer cells. This feature obviates the need for time-consuming plaque assays to isolate, screen and select viral clones prior to large-scale AdV propagation and purification. Recently, the versatility of this method was increased by the development of an AdEasy-based system to produce first- and second-generation AdVs with tropism for CAR- or CD46-positive cells (38). In this chapter, we exploit this new AdV assembly system for detailing the construction and characterization of second-generation fiber-modified AdVs containing *TALEN* expression units. The same methods can be applied to generate AdVs with other genotypes.

1- Insert each *TALEN* ORF into an AdEasy-compatible pAdShuttle expression plasmid by using conventional recombinant DNA procedures (**Fig. 1**) (39). In the present work, the *TALEN* ORFs in 1383.pVAX.AAVS1.TALEN.L-94 and 1384.pVAX.AAVS1.TALEN.R-95 (32) were inserted into pAdSh.PGK.SV40 yielding expression plasmids pAdSh.PGK.TALEN-L^{S1} and pAdSh.PGK.TALEN-R^{S1}, respectively. Of note, the decision on which promoter to use is an important one and should be done having in sight which cell type(s) are intended as target cells. For instance, it is known that the ubiquitously used cytomegalovirus *immediate-early* gene promoter, albeit strong in many cell types, underperforms or suffers from silencing phenomena in others. Among the latter are those with substantial scientific and therapeutic relevance such as muscle cells and embryonic stems cells. Therefore, to broaden the use of transient AdV-mediated TALEN expression in human cells, it might be advisable to select cellular gene regulatory elements. Generally, valuable candidates are represented by the promoter/enhancer elements derived from the human genes *EEF1A1*, *UBC* and *PGK1* (a.k.a. *EF1α*, *ubiquin C* and *PGK*, respectively).

2- Digest 4 µg of TALEN-encoding pAdShuttle plasmids (e.g. pAdSh.PGK.TALEN-L^{S1} and pAdSh.PGK.TALEN-R^{S1}) with the restriction enzyme PmeI (or MssI isoschizomer) in a total volume of 40 µl. Check the completeness of the digestion by subjecting a 10-µl sample to agarose gel electrophoreses.

3- Precipitate the digested DNA by adding 20 µl of demineralized water, 25 µg of tRNA "carrier" molecules plus 2.5 volumes of ice-cold absolute ethanol, homogenize and centrifuge the mixtures at 20,000 ×g for 30 minutes at 4ºC.

4- Dissolve the recovered dried DNA pellet in 30 µl of demineralized water and use 10 µl to transform, in an ice-cold 1-mm cuvette, a 90-µl suspension of electro-competent BJ5183 cells containing the pAdEasy "backbone" of choice. In this chapter, to package the TALEN expression units in second-generation fiber-modified AdV particles, we used BJ5183[pAdEasy-2.50] cells (38). BioRad Gene Pulser electroporation settings: 1.25 V, 200 Ω and 25 µFD with a time constant of about 4 seconds.
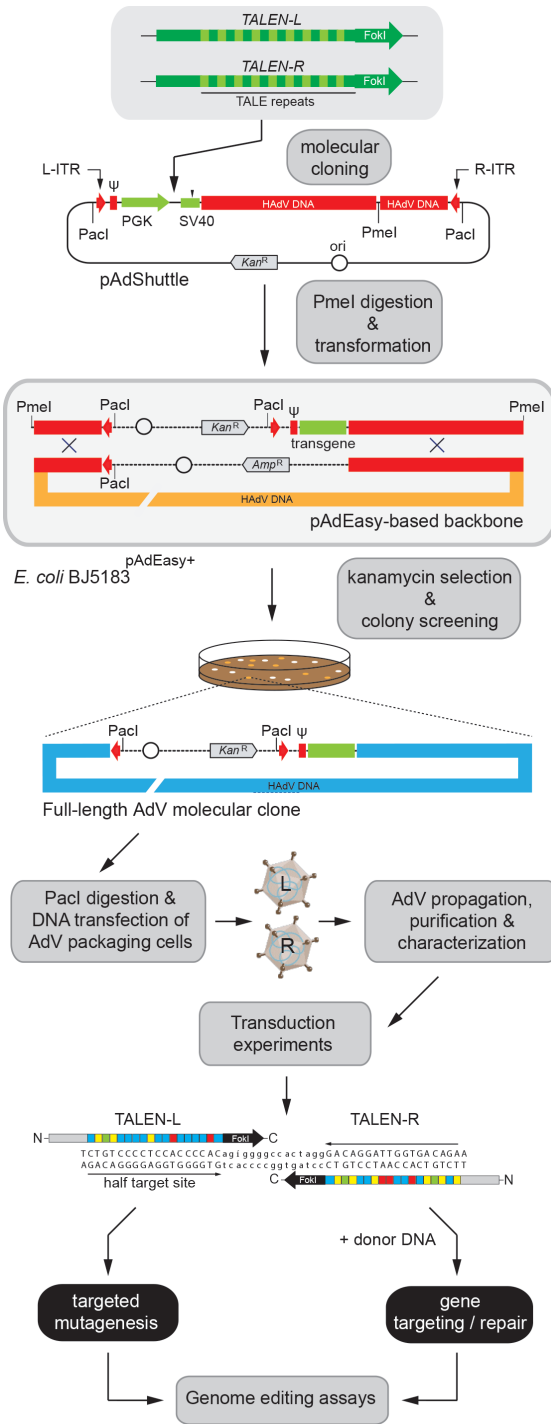
Figure 1. Outline of the AdV-based TALEN packaging and delivery system. The ORFs coding for a functional heterodimeric TALEN pair (*TALEN-L*/*TALEN-R*) are individually cloned into the multiple cloning site of an AdEasy "shuttle" plasmid (pAdShuttle) containing the transcriptional regulatory elements of choice (e.g. the human *PGK1* gene promoter [PGK] and the simian virus 40 polyadenylation signal [SV40]). L-ITR and R-ITR, HAdV "left" and "right" inverted terminal repeats, respectively; Ψ, HAdV packaging signal; TALE repeats, transcription activator-like effector region coding for the sequence-specific DNA binding domain; TALEN-L and TALEN-R, "left" and "right" TALE nuclease, respectively; FokI, DNA portion encoding the non-specific nuclease domain of the homodimer-forming type IIS restriction enzyme FokI; *Kan*R, kanamycin-resistance gene; *ori*, prokaryotic origin of replication. The resulting constructs are treated with PmeI and are transformed into recombinogenic *E. coli* strain BJ5183[pAdEasy+] harboring a pAdEasy-derived "backbone" construct of choice. In the current work we deployed BJ5183[pAdEasy-2.50] cells to generate *E1*- and *E2A*-deleted (i.e. second-generation) AdVs displaying particle-retargeting fiber motifs from HAdV serotype 50. HR at DNA sequences common to "shuttle" and "backbone" constructs (labeled in the red) results in the assembly of full-length AdV molecular clones. These plasmids are isolated by screening individual colonies grown in agar plates containing kanamycin. Transfection of AdV packaging cell lines with PacI-digested AdV molecular clones leads to the assembly of AdV particles containing TALEN-encoding expression units. The rescued vector particles can subsequently be propagated to high-titers in producer cells and purified from cellular components. The resulting stocks of AdV particles are then subjected to assays to establishing the quantity and integrity of their packaged recombinant genomes. After this final characterization, AdV preparations are ready for genome editing experiments based on transducing and expressing the functional TALEN pair (TALEN-L/TALEN-R) in human cells.

5- Add 400 µl of LB broth medium without antibiotics, transfer the bacteria to a sterilized Eppendorf tube and shake for 45-60 minutes at 30ºC.

6- Plate the transformed cells onto two LB agar plates containing 50 µg/ml of kanamycin sulfate and, after an overnight incubation period at 30ºC, select and grow small-sized colonies for 16-20 hours at 30ºC in the presence of 50 µg/ml of kanamycin sulfate.

7- Isolate the plasmids by using a conventional alkaline lysis DNA extraction protocol and, on the basis of restriction enzyme analyses, select clones containing full-length AdV genomes for large-scale plasmid DNA purification. To this end, any commercially available DNA extraction procedure based on ion-exchange chromatography columns can be deployed. Optional extra step: transform into and purify from a non-recombinogenic *E. coli* strain (e.g. DH5α) the full-length AdV molecular clones.

8- Determine the concentration and purity (A_{260}/A_{280}) of the resulting large-scale DNA preparations by spectrophotometry and confirm their integrity by restriction enzyme fragment length analyses. If warranted, complement these data through transgene-directed DNA sequencing analyses.

RESCUE AND PROPAGATION OF ADVS

Materials
1. Full-length AdV molecular clones (here pAdV^{Δ2}P.TALEN-L^{S1}.F^{50} and pAdV^{Δ2}P.TALEN-R^{S1}.F^{50}).
2. PacI restriction enzyme (New England Biolabs)
3. 1 mg/ml 25-kDa linear polyethylenimine (PEI; Polysciences) in H$_2$O
4. 0.5 M NaCl (Merck)
5. 5 ml Polystyrene Round-Bottom tubes (BD Falcon)
6. *E1*- and *E2A*-complementing AdV packaging cell line (here PER.E2A cells)
7. Dulbecco's modified Eagle's medium (DMEM; Life Technologies)
8. Non-heat inactivated fetal bovine serum (FBS; Life Technologies)
9. 1M MgCl$_2$ (Sigma-Aldrich)
10. 50 ml blue screw cap conical bottom tubes (Greiner Bio-One)
11. 0.05% trypsin-ethylenediaminetetraacetic (EDTA) solution (Life Technologies)
12. Phosphate-buffered saline (PBS) pH=7.4
13. 6-well plates, 75- and 175-cm² cell culture flasks (all from Greiner Bio-One)
14. 0.45 µm pore-sized filters (Pall Life Sciences)
15. Liquid N$_2$ and water baths

Method

In brief, full-length AdV molecular clones (e.g. pAdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and pAdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50}) are first linearized with PacI (**Fig. 2A**). Next, the PacI-treated DNA is introduced into AdV packaging cells by a polycation-based DNA transfection protocol. The restriction enzyme treatment removes the prokaryotic DNA sequences and positions the ITR-embedded HAdV origins of replication close to DNA termini necessary for the initiation of HAdV-dependent replication. The rescued genome-containing AdV particles are subsequently amplified to high-titers by serial propagation on increasing numbers of producer cells.

1- Digest 6.25 µg of each full-length AdV molecular clone with PacI in a total volume of 30 µl, after which heat-inactivate the restriction enzyme by a 20-minute incubation period at 65ºC.

2- Seed PER.E2A cells (40) or any another *E1-* and *E2A*-complementing cell line in wells of 6-well plates. In the case of PER.E2A cells, seed 1.25×10^6 cells in DMEM containing 10% FBS and 10 mM MgCl$_2$. Incubate overnight at 39ºC in a 10% CO$_2$ atmosphere. The mutant *E2A* gene in PER.E2A cells confers a thermo-sensitive phenotype to the encoded DNA-binding protein (DBP). Therefore, these cells are normally cultured at 39ºC (non-permissive temperature) to avoid DBP's cytotoxicity, whereas at the time of vector production, they are shifted to 34ºC (permissive temperature) to allow proper folding of the protein. Clearly, to rescue first-generation AdVs with conventional or retargeting fibers, *E1*-complementing cell lines such as 293 (41), 911 (42), or PER.C6 (27) suffice.

3- Add the 30-µl PacI digestion mixture to 190 µl of 150 mM NaCl in a 5 ml polystyrene round-bottom tube. In parallel, dilute 18.8 µl of a 1 mg/ml PEI solution in 150 mM NaCl to a final volume of 220 µl.

4- Add the PEI to the DNA solution, vortex vigorously for approximately 10 seconds and incubate for 20 minutes at room temperature to let the DNA-polycation complexes form. Next, dispense dropwise the colloidal suspension into the wells containing the AdV packaging cells.

5- After overnight incubation at 39ºC in a 10% CO$_2$ atmosphere, remove the transfection mixtures, add regular culture medium and incubate the cells at the permissive temperature of 34ºC in a 10% CO$_2$ atmosphere. Monitor the monolayers for the appearance of viral plaques and, ultimately, complete cytopathic effect (CPE) as defined by over 90% detached producer cells. Of note, in case the transfected cells become over-confluent stalling the development of full CPE, sub-culture them into a 75-cm^2 flask taking care to beforehand remove the trypsin by pelleting and re-suspending the produced cells in regular fresh medium.

6- Release the rescued AdV particles from the producer cells by subjecting the collected cell suspension in medium to three cycles of freezing and thawing. To this end, liquid N$_2$ and 37ºC water baths can be used.

7- Remove the cellular debris by centrifugation at 1,000 ×*g* for 10 minutes in a swing-out rotor and add supernatant fractions to 75-cm² culture flasks containing newly seeded producer cells at about 70-80% confluence. Of note, for optimal AdV propagation, it is advisable adjusting the inocula fractions to result in complete CPE of the producer cells within 2 to 3 days post-infection. As an optional extra step, the recovered supernatants can be further clarified by filtration through 0.45 µm pore-sized filters. To generate a research-scale AdV preparation, a minimum of 1 and a maximum of 3 propagation rounds are normally sufficient to yield enough clarified AdV supernatant material ("master virus seed stock") to use in at least one large-scale infection round comprising 18-20 175-cm² culture flasks containing producer cells at 70-80% confluence (~2×10⁷ cells/flask).

8- Collect the producer cells in sterile 50 ml blue screw cap conical bottom tubes and centrifuge them at 2,000 ×*g* for 10 minutes in a swing-out rotor. Discard the supernatants, resuspend the producer cells in 40 ml of PBS and divide the cell suspensions in two sterile 50 ml blue screw cap conical bottom tubes. Move forward to the concentration and purification of the AdV particles or store the material at -80ºC.

CONCENTRATION AND PURIFICATION OF ADV PARTICLES

Materials
1. Ultracentrifuge (here Beckman Coulter LE-80K)
2. SW28 rotor (Beckman Coulter) or similar
3. VTI65.1 rotor (Beckman Coulter) or similar
4. Open-top polyallomer centrifuge tubes, 25×89 mm (Beckman Coulter)
5. Quick-Seal® polyallomer centrifuge tubes, 16×76 mm (Beckman Coulter)
6. Tube Sealer 342420 (Beckman Coulter) or similar apparatus
7. Amicon Ultra-15 Centrifugal Filter Devices with 100,000 MWCO (Millipore) or similar device
8. 5% (w/v) sodium deoxycholate (DOC; Merck) in demineralized water
9. 1.24 g/ml, 1.33 g/ml and 1.4 g/ml cesium chloride (CsCl; Merck) solutions in Tris-HCl pH 7.5 with 5% glycerol
10. 10 mg/ml DNaseI (grade II; Roche Applied Science)
11. 20G needles (BD microlance 3) and 5-ml syringes (BD plastipak)
12. Storage buffer (20 mM Tris-HCl pH 8.0; 25 mM NaCl and 5% glycerol)

Methods

In brief, the AdV particles are first released from producer cells by a mild detergent treatment. Next, they are concentrated and purified by two consecutive ultracentrifugation steps entailing CsCl block and continuous isopycnic density gradients. Banded genome-containing AdV particles (**Fig. 2B**) are subsequently subjected to a de-salting buffer exchange protocol based on ultrafiltration.

1- Add to 20-ml suspensions of producer cells in PBS the detergent DOC at a final concentration of 0.5% and incubate the mixtures at 37ºC for 30 minutes. Gently invert the cell lysates every 5-10 minutes for a thorough homogenization of the resulting viscous solution.

2- Add $MgCl_2$ and DNaseI at final concentrations of 40 mM and 8.7 µg/ml, respectively. Subsequently, incubate the mixtures at 37ºC for 30 minutes while gently inverting the cell lysates every 5-10 minutes for a thorough homogenization.

3- Centrifuge the producer cell lysates for 10 minutes at 2000 ×*g*. Next, recover the supernatants and centrifuge them for 10 minutes at 3000 ×*g*. The resulting clarified cell lysates are subsequently subjected to ultracentrifugation through block and continuous CsCl density gradients as follows.

4- Add, in a gentle manner, 20-22 ml of clarified supernatant material on top of block gradients consisting of 5 ml and 10 ml CsCl tires with densities of 1.4 g/ml and 1.24 g/ml, respectively. Open-top polyallomer centrifuge tubes with a 38-ml capacity can be used for this purpose.

5- Place the loaded centrifuge tubes in a SW28 swing-out rotor (or similar) and centrifuge them at 80,000 ×*g* for 2 hours at 10ºC. Of note, for proper counterbalancing during ultracentrifugation, the weight differences between tubes placed at opposite rotor positions should not exceed 0.02 grams.

6- Recover the AdV particles concentrated at the border between the two CsCl densities by pricking the centrifuge tubes with a 20G needle coupled to 5-ml syringe.

7- Transfer the recovered AdV material into 13.5-ml polyallomer Quick-Seal® tubes and fill them up completely with 1.33 g/ml CsCl solution. Next, heat-seal the tubes by using a pre-heated Tube Sealer 342420 (or similar device). In alternative to heat-sealing, use Opti-Seal® polyallomer centrifuge tubes instead. Of note, for proper counterbalancing during ultracentrifugation, the weight differences between tubes placed at opposite rotor positions should not exceed 0.02 grams after sealing.

8- Place the loaded and sealed centrifuge tubes in a VTi 65.1 vertical-tube titanium rotor (or similar) and centrifuge them at 300,000 ×*g* for 16 hours at 10ºC to subject the AdV particles to a self-generating continuous buoyant density gradient. Of note, disable the break option of the ultracentrifuge prior to stopping these runs.

9- Use a needle to create a hole at the top of the sealed centrifuge tubes after which retrieve the banded AdV particles (**Fig. 2B**) by piercing the tubes with 20G needles coupled to 5-ml syringes.
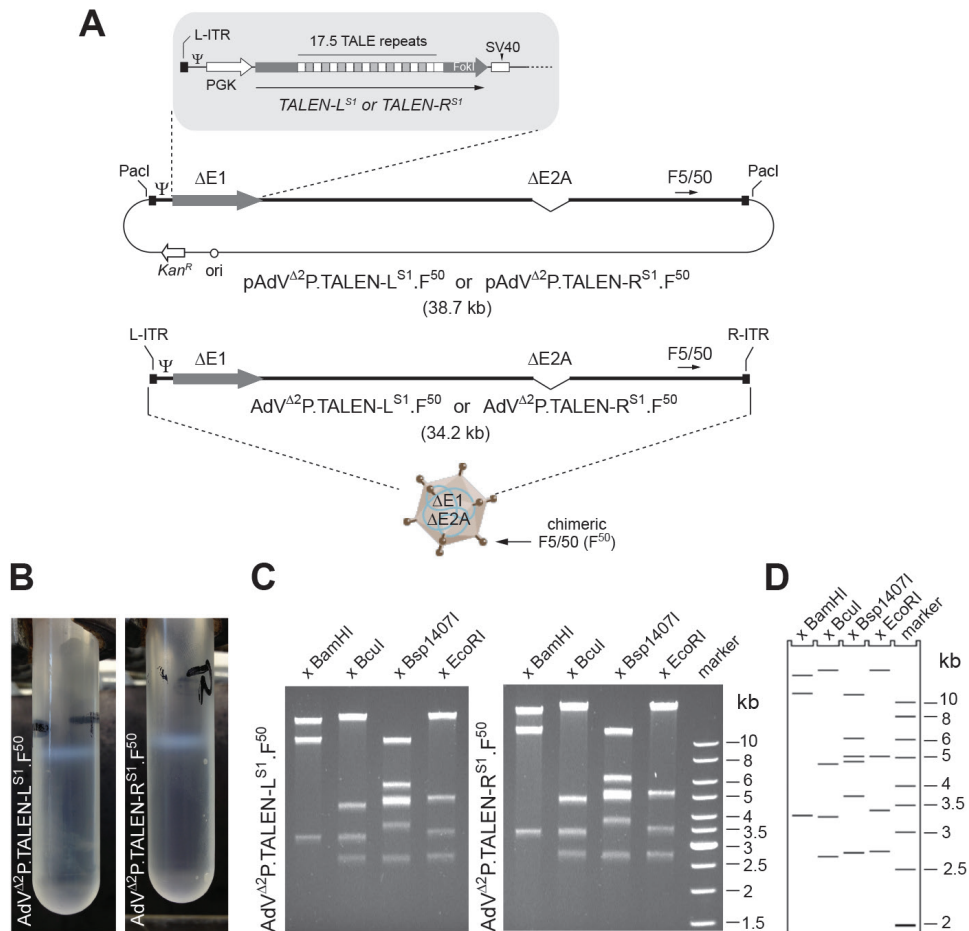
Figure 2. Analysis of recombinant genomes packaged in purified AdV particles. **(A)** DNA structures of second-generation (AdV$^{\Delta2}$) fiber-modified (F^{50}) adenoviral vectors AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} drawn in relation to those of their respective molecular clones pAdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and pAdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}. The thick and thin lines correspond to the AdV and plasmid backbone sequences, respectively. F5/50, ORF coding for a chimeric fiber consisting of basal shaft domains from HAdV serotype 5 fused to apical shaft and knob motifs from HAdV serotype 50. For an explanation of the other elements see the legend of Fig. 1. These recombinants contain an expression unit coding for the *AAVS1*-specific designer nucleases TALEN-L^{S1} or TALEN-R^{S1}. The regulatory elements consist of the *PGK* promoter and the SV40 polyadenylation signal. The central portions of these *TALEN* ORFs encode an array of 17.5 repeats responsible for targeting the non-specific FokI nuclease domains to *AAVS1* to form an operative TALEN pair. **(B)** Aspect of concentrated genome-containing AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} particles after isopycnic CsCl density-gradient ultracentrifugation. **(C)** Characterization of DNA isolated from CsCl-purified AdV particles by restriction fragment length analysis. Agarose gel electrophoresis of AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} genomes (left and right panels, respectively) digested with BamHI, BcuI, Bsp1407I or EcoRI. Marker, GeneRuler DNA Ladder Mix molecular weight marker (Fermentas). **(D)** Restriction patterns of AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} DNA generated *in silico* with the aid of Gene Construction Kit (version 2.5) software (Textco BioSoftware, Inc.).

10- Transfer the recovered AdV material into Amicon Ultra-15 Centrifugal Filter Devices with a nominal molecular weight cut-off of 100 kDa previously rinsed with storage buffer.

11- Fill-up the filter devices with storage buffer to their maximum 15 ml capacity and exchange the CsCl solution for the physiological storage buffer by carrying out a minimum of five consecutive cycles of centrifugation and addition of large volumes of fresh storage buffer. Of note, avoid drying the filter membrane by overextending the centrifugation step(s). As a guideline for the first centrifugation round, apply a centrifugal force of 2,000 ×*g* for 45 seconds at room temperature. For subsequent rounds adapt the centrifugation time depending on the retentate and flow-through volumes corresponding to each AdV preparation.

12- After the last cycle, recover the retentate (1.5-2.0 ml) into sterilized Eppendorf tubes or 5 ml polystyrene round-bottom tubes and rinse well the filter membranes 2-3 times with 200 µl of storage buffer. Pool this material into the bulk material, mix and make aliquots. Store the aliquots at -80ºC until their use.

CHARACTERIZATION OF ADV PREPARATIONS

The characterization of purified stocks of TALEN-encoding AdVs includes determining the integrity and amounts of packaged recombinant genomes (sections **5.1** and **5.2**, respectively) as well as validating the functionality of their expression units and encoded products (section **5.3**). After being extracted from AdV particles, the integrity of the vector genomes can be routinely examined by conventional restriction enzyme fragment length analyses (**Fig. 2C** and **Fig. 2D**). If necessary, these data can be complemented by more detailed analyses based on DNA sequencing of selected AdV DNA portions (e.g. *TALEN* ORFs). In addition, recombinant AdV genomes in purified preparations can be quantified by qPCR or, alternatively, by a quicker procedure based on disrupting AdV capsids with an ionic detergent and directly measuring the released vector DNA content with the aid of a dsDNA-specific fluorescent dye. Here we deploy such a method, adapted from Murakami and McCaman (43), to determine AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} titers in terms of genome-containing viral particles per ml (VP/ml). This method allows for accurate and rapid titer measurements by exploiting the sensitivity and wide dynamic range of the dsDNA-specific dye PicoGreen (**Fig. 3**). Finally, the functionality of AdV-delivered *TALEN* transcriptional units can be established by exploiting heterologous protein tags and standard western blot techniques (**Fig. 4A**) as well as surrogate NHEJ detecting assays (**Fig. 4B**) based on mismatch-sensitive endonucleases (e.g. T7 endonuclease I or CEL-1).
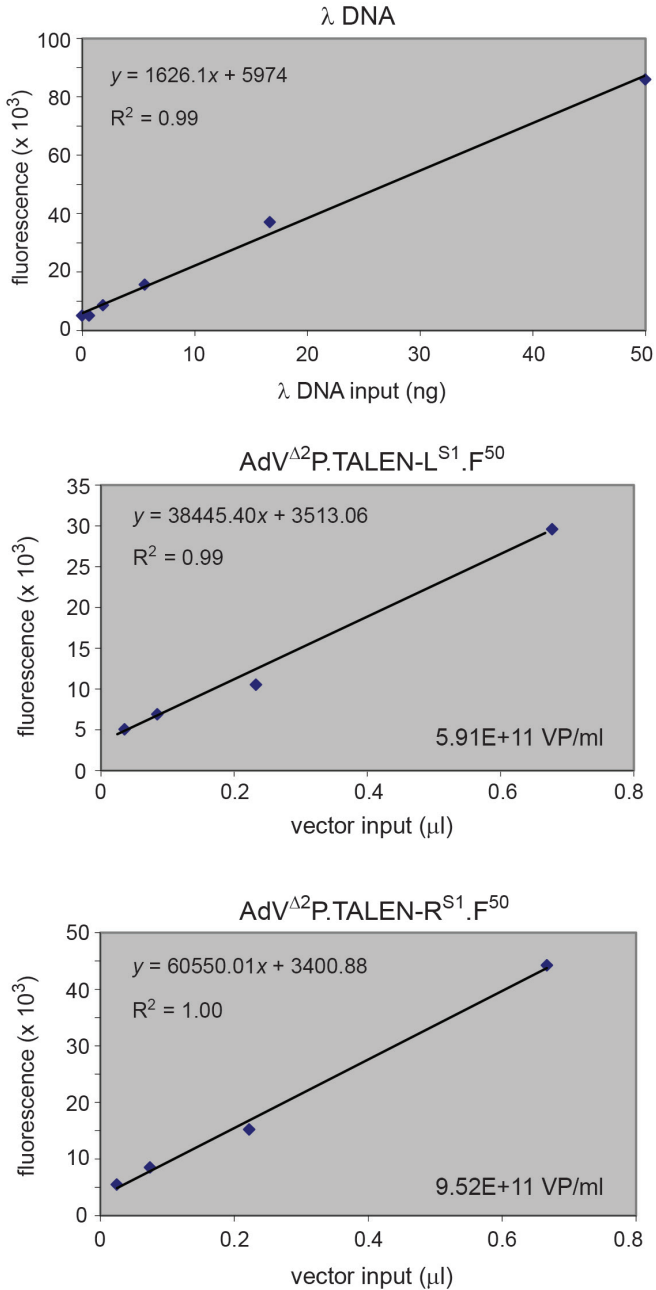
Figure 3. Fluorometric assay based on the double-stranded DNA dye PicoGreen for measuring AdV titers in terms of genome-containing viral particles per milliliter (VP/ml). Representative linear regression curves corresponding to a λ DNA standard curve and to DNA from CsCl-purified AdV stocks (i.e. AdV^{Δ2}P.TALEN-L^{S1}.F^{50} and AdV^{Δ2}P.TALEN-R^{S1}.F^{50}) are depicted. The resulting AdV physical particle titers for each stock are also shown.
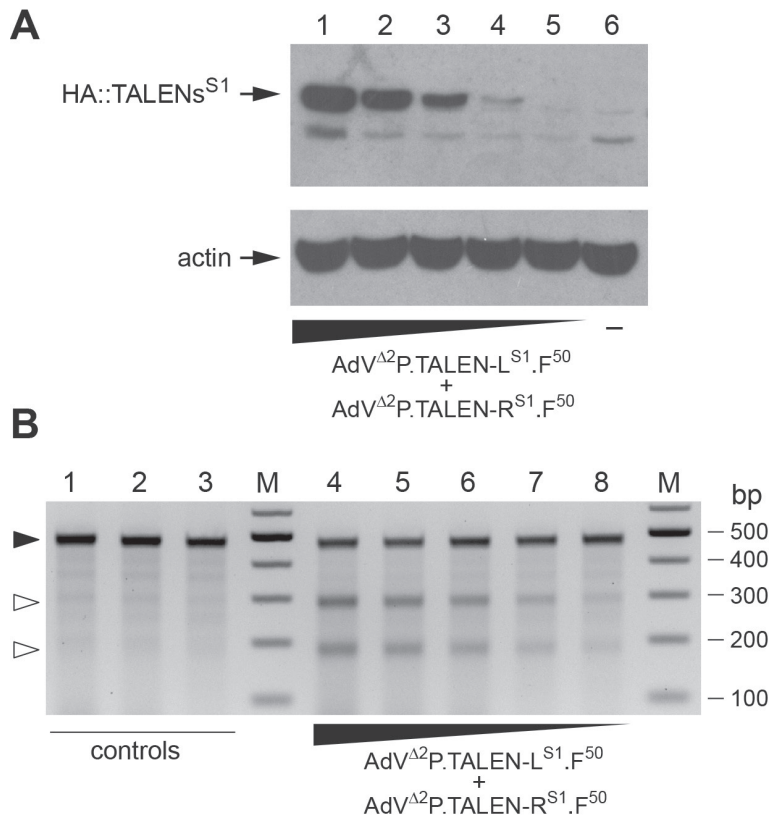
Figure 4. Transgene expression and targeted mutagenesis assays for the validation of TALEN-encoding AdV preparations. (**A**) Influenza hemagglutinin tag-specific Western blot analyses on protein lysates of human myoblasts co-transduced with 1:1 mixtures of $AdV^{\Delta 2}P.TALEN-L^{S1}.F^{50}$ and $AdV^{\Delta 2}P.TALEN-R^{S1}.F^{5}$ at total MOIs of 15×10^3, 10×10^3, 5×10^3, 2.5×10^3 and 1×10^3 VP/cell (lanes 1 through 5, respectively) or mock-transduced (lane 6). A parallel Western blot deploying an anti-actin mouse monoclonal antibody (clone: C4) provided for a loading control. (**B**) T7 endonuclease I genotyping assay to detect site-specific DSBs in target cells exposed for three days to TALEN-encoding AdVs. Human myoblasts were co-transduced with 1:1 mixtures of $AdV^{\Delta 2}P.TALEN-L^{S1}.F^{50}$ and $AdV^{\Delta 2}P.TALEN-R^{S1}.F^{5}$ at total MOIs of 15×10^3, 10×10^3, 5×10^3, 2.5×10^3 and 1×10^3 VP/cell (lanes 4 through 8, respectively). Controls were provided by mock-transduced human myoblasts (lane 3) or by human myoblasts exposed exclusively to 7.5×10^3 VP/cell of $AdV^{\Delta 2}P.TALEN-L^{S1}.F^{50}$ (lane 1) or to 7.5×10^3 VP/cell of $AdV^{\Delta 2}P.TALEN-R^{S1}.F^{50}$, (lane 2). Lanes M, GeneRuler DNA Ladder Mix molecular weight marker (Fermentas). PCR products spanning the *AAVS1* target site of control human myoblasts or of human myoblasts transduced with different doses of $AdV^{\Delta 2}P.TALEN-L^{S1}.F^{50}$ and $AdV^{\Delta 2}P.TALEN-R^{S1}.F^{50}$ are denatured and re-annealed. Heteroduplexes with indels resulting from NHEJ repair of DSBs at *AAVS1* bear mismatched base pairs that can be recognized and cleaved by T7 endonuclease I yielding low-molecular-weight DNA fragments whose positions are indicated by open arrowheads.

Structural analysis of packaged AdV genomes

Materials

1. 10× DNaseI buffer (130 mM Tris-HCl pH 7.5; 1.2 mM $CaCl_2$ and 50 mM $MgCl_2$)
2. 10 mg/ml DNaseI (Roche)
3. 0.5 mM EDTA pH 8
4. 20 mg/ml Proteinase K (Thermo Scientific)
5. 10% (w/v) SDS
6. Jetsorb Gel Extraction Kit (Genomed) or similar system
7. Selected restriction enzymes
8. Agarose gel electrophoresis reagents and apparatus

Methods

1- Add to a 50-µl aliquot of a CsCl-purified AdV preparation, 12 µl of 10× DNaseI buffer, 8 µl of 10 mg/ml DNaseI and 50 µl of demineralized water. Incubate the mixture for 30 minutes at 37ºC.

2- Inactivate the DNaseI by adding 2.4 µl of 0.5 mM EDTA pH 8, 6.0 µl of 10% SDS and 1.5 µl of 20 mg/ml proteinase K. Incubate the mixture for 1 hour at 55ºC.

3- Add 390 µl of buffer A1 (Jetsorb Gel Extraction Kit) and 10 µl of silica beads (Jetsorb Gel Extraction Kit). Incubate for 20 minutes and homogenize every 3-5 minutes by inverting the samples several times.

4- Centrifuge for 30 seconds at 20,000 ×*g*, discard the supernatant and gently resuspend the bead pellets in 400 µl of buffer A1.

5- Centrifuge for 30 seconds at 20,000 ×*g*, discard the supernatant and gently resuspend the bead pellets in 400 µf of buffer A2 (Jetsorb Gel Extraction Kit). Repeat this washing step and remove the supernatants as thoroughly as possible.

6- Air-dry the bead pellets until they become white. Of note, avoid over-drying the pellets to prevent difficulties in eluting the bound DNA.

7- Elute the DNA by adding 30 µl of nuclease-free demineralized water and incubate for 7 minutes at 55ºC. Increase the DNA elution by tapping the tubes approximately every 2 minutes.

8- Centrifuge the samples at 20,000 ×*g* for 30 seconds and carefully collect the supernatants containing the eluted AdV genomes. Of note, avoid carrying over silica beads as these can inhibit subsequent enzymatic reactions.

9- Perform restriction enzyme fragment length analysis to establish the integrity of the packaged AdV genomes (**Fig. 2C** and **Fig. 2D**) and, if warranted, carry out transgene-directed DNA sequencing.

Determination of AdV physical particle titers
Materials
1. Storage buffer (20 mM Tris-HCl pH 8.0; 25 mM NaCl and 5% glycerol)
2. Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies)
3. White 96-well microplate OptiPlate-96 (Perkin Elmer)
4. TE buffer (10mM Tris-HCl and 1 mM EDTA pH 7.5)
5. Spectrofluorometer plate reader (here Wallace 1420 VICTOR 3 multilabel plate reader, Perkin Elmer)

Methods
1- Dilute 5 μl of each AdV stock in 45 μl of storage buffer supplemented with 0.05% SDS. Next, prepare four 3-fold serial dilutions by mixing 15 μl of the vector material with 30 μl of storage buffer containing 0.05% SDS.

2- Incubate the dilution series for 5 minutes at 65ºC, after which pipet 20 μl of each of these samples into wells of white 96-well microplates.

3- Prepare a working solution of the fluorescent dye by diluting the stock of Quant-iT PicoGreen dsDNA 360 fold in TE buffer. Next, add 180 μl of this working solution to each of the dilutions. Incubate for 5 minutes in the dark at room temperature.

4- In parallel, set-up a calibration curve by serially diluting a stock of bacteriophage λ DNA (100 μg/ml) in storage buffer containing 0.05 % SDS to the final concentrations: 50, 16.7, 5.6, 1.9 and 0.6 ng/μl.

5- Measure the various DNA concentrations by deploying a Wallace 1420 VICTOR 3 multilabel plate reader (or similar device) set to excitation and emission wavelengths of 485 nm and 535 nm, respectively.

6- Generate linear regression curves for each of the measured data sets confirming that the corresponding coefficients of determination (R^2) lie between 0.98 and 1.0 (**Fig. 3**). Next, through the λ DNA standard curve correlating fluorescence signals to ng of DNA, determine for each AdV stock the amount of DNA (in ng) per μl of sample.

7- Calculate the molecular mass corresponding to each AdV DNA molecule by applying the relationship: Genome size (bp) × 660 (Da/bp) = Da (gmol^{-1}).
e.g. AdV^{Δ2}P.TALEN-L^{S1}.F^{50} DNA = 34,155 bp × 660 Da/bp = $2.25×10^7$ gmol^{-1}.

8- Divide the resulting values by the Avogadro's number (i.e. $6.022×10^{23}$) to obtain the weight of each AdV DNA molecule in grams and subsequently convert these values to ng; e.g. $2.25×10^7$ gmol^{-1}/$6.022×10^{23}$ mol = $3.74×10^{-17}$ g → $3.74×10^{-8}$ ng

9- Determine the number of AdV genomes per ng; e.g. 1/$3.74×10^{-8}$ ng = $2.67×10^7$ genomes

10- Use the values obtained in step **6** to calculate the total number of genomes per μl; e.g. $2.67×10^7$ genomes/ng × 22.1 ng/μl = $5.91×10^8$ genomes/μl → $5.91×10^{11}$ genomes/ml or genome-containing viral particles per ml i.e. VP/ml.

Functional validation of AdV-delivered TALENs

Materials

1. Cell lysis solution for genomic DNA extraction (100 mM Tris-HCl pH 8.5; 5 mM EDTA; 0.2% SDS and 200 mM NaCl)
2. TE-saturated phenol:chloroform:isoamylalcohol (25:24:1) solution (J.T. Baker)
3. Chloroform (Merck)
4. 7.5 M ammonium acetate (Merck)
5. Absolute ethanol (J.T. Baker)
6. 10 mM TE buffer pH 8.0
7. 10 mg/ml RNase A (Thermo Scientific)
8. 5U/µl GoTaq DNA polymerase (Promega)
9. 5× Colorless GoTaq Flexi buffer (Promega)
10. 40 mM deoxynucleotide (dNTP) solution mix (New England Biolabs) and 10 µM selected primer pair (Eurofin)
11. PCR thermocycler (here DNA Engine Tetrad 2 Peltier Thermal Cycler [Bio-Rad])
12. Agarose gel electrophoresis reagents and apparatus
13. 5× loading buffer
14. 10 U/µl T7 endonuclease I and 10× NEBuffer 2 (both from New England Biolabs)
15. DNA molecular-weight marker e.g. Gene Ruler DNA Ladder Mix (Thermo Scientific)
16. Bio-Rad Image Lab 4.1 (Bio-Rad) or ImageJ (National Institutes of Health, USA) software

Methods

The below-specified transduction experiments serve simply as an example to provide general guidelines for validating TALEN-encoding AdVs. Obviously, types and amounts of human target cells as well as vector genotypes and doses are at the discretion of the investigator.

1- One day prior transduction, seed $2.0×10^5$ human myoblasts in wells of a 24-well plate and incubate the cells overnight at 37ºC in a 5% CO_2 atmosphere.

2- Co-transduce the target cells with 1:1 mixtures of AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50}. Preferably, to determine limiting doses and TALEN activity plateauing levels, apply a broad vector dose range onto the target cells. Negative controls can be provided by mock-transduced cells as well as by cells exposed at high doses to each AdV individually.

3- At 3 days post-infection, harvest the control and the co-transduced cell cultures and divide them in samples for protein expression and target site genotyping analysis. The assay for the former analyses has been detailed elsewhere (32), whereas that for the latter is as follows.

4- Initiate the total cellular DNA extraction procedure by incubating the target cells overnight at 55ºC in 500 µl of lysis buffer containing freshly added proteinase K at a final concentration

of 100 ng/ml. Next, subject the DNA to two rounds of phenol:chloroform:isoamylalcohol (25:24:1) extraction and to one round of chloroform extraction to remove traces of phenol. Next, add half volume of 7.5 M ammonium acetate and 2.5 volumes of ice-cold ethanol and precipitate the DNA by centrifugation at 20,000 ×g for 30 minutes at 4ºC. Wash the DNA pellets with 70% ethanol, after which air-dry and resuspend the DNA in 100 µl TE buffer pH 8.0 containing 100 µg/ml of DNase-free RNaseI.

6- Add 2 µl of the purified DNA to 50-µl PCR mixtures containing 0.4 µM of *AAVS1*-specific primers (5'-TTCGGGTCACCTCTCACTCC-3' and 5'-GGCTCCATCGTAAGCAAACC-3'), 0.1 mM of each dNTP, 1× Colorless GoTaq Flexi buffer, 1 mM $MgCl_2$ and 2.5 U GoTaq DNA polymerase.

7- Apply the following PCR cycling conditions: Initial denaturation for 5 minutes for 95ºC, followed by 40 cycles of 30 seconds at 95ºC, 30 seconds at 61ºC and 30 seconds at 72ºC. Terminate the reactions by an elongation period of 5 minutes at 72ºC.

8- Denature and re-anneal 20-µl amplicon samples by using a regular thermocycler machine and the following program:

95°C 10 min. 95°C to 85°C (-2°C/sec.)	85°C 1 min. 85°C to 75°C (-0.3°C/sec.)
75°C 1 min. 75°C to 65°C (-0.3°C/sec.)	65°C 1 min. 65°C to 55°C (-0.3°C/sec.)
55°C 1 min. 55°C to 45°C (-0.3°C/sec.)	45°C 1 min. 45°C to 35°C (-0.3°C/sec.)
35°C 1 min. 35°C to 25°C (-0.3°C/sec.)	25°C 1 min. 4°C hold
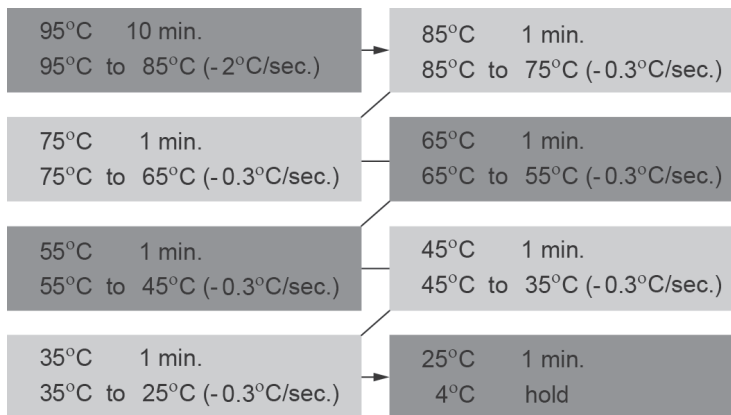
9- Incubate the resulting DNA species in 25-µl reactions containing 5 U of T7 endonuclease I and 1× NEBuffer 2 for 20 minutes at 37ºC.

10- Stop the reactions by adding loading sample buffer containing SDS and immediately subject the DNA to electrophoresis through a 1.5% agarose gel in 1× Tris-Acetate-EDTA buffer.

11- Visualize by in-gel staining with a DNA dye (e.g. ethidium bromide) intact and digested amplicons resulting from NHEJ-mediated formation of small insertions and deletions (indels) at the TALEN target sequences.

12- Estimate the frequency of indel formation by densitometry of electrophoresis-resolved DNA molecules. To this end, commercially and publicly available software packages such as Bio-Rad Image Lab 4.1 and ImageJ, respectively, can be deployed.

ACKNOWLEDGMENTS

4

REFERENCES

1. Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773-782.

2. Corrigan-Curay, J., Cohen-Haguenauer, O., O'Reilly, M., Ross, S.R., Fan, H., Rosenberg, N., Somia, N., King, N., Friedmann, T., Dunbar, C. *et al.* (2012) Challenges in vector and trial design using retroviral vectors for long-term gene correction in hematopoietic stem cell gene therapy. *Mol Ther*, **20**, 1084-1094.

3. Chuah, M.K. and Vandendriessche, T. (2013) Optimizing delivery and expression of designer nucleases for genome engineering. *Hum Gene Ther Methods*, **24**, 329-332.

4. Gaj, T., Gersbach, C.A. and Barbas, C.F., 3rd. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, **31**, 397-405.

5. Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol*, **29**, 816-823.

6. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods*, **8**, 765-770.

7. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, **31**, 822-826.

8. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*, **31**, 839-843.

9. Cradick, T.J., Fine, E.J., Antico, C.J. and Bao, G. (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*, **41**, 9584-9592.

10. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, **39**, 9283-9293.

11. Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., Santiago, Y., Vincent, A.I., Meng, X., Zhang, L. *et al.* (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol*, **29**, 695-696.

12. Joung, J.K. and Sander, J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, **14**, 49-55.

13. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.

14. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.

15. Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol*, **31**, 76-81.

16. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*, **30**, 460-465.

17. Gonçalves, M.A. and de Vries, A.A. (2006) Adenovirus: from foe to friend. *Rev Med Virol*, **16**, 167-186.

18. Kaufmann, J.K. and Nettelbeck, D.M. (2012) Virus chimeras for gene therapy, vaccination, and oncolysis: adenoviruses and beyond. *Trends in molecular medicine*, **18**, 365-376.

19. Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H. and Ward, T.G. (1953) Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med*, **84**, 570-573.

20. Arnberg, N. (2012) Adenovirus receptors: implications for targeting of viral vectors. *Trends in pharmacological sciences*, **33**, 442-448.

21. Bergelson, J.M., Cunningham, J.A., Droguett, G., Kurt-Jones, E.A., Krithivas, A., Hong, J.S., Horwitz, M.S., Crowell, R.L. and Finberg, R.W. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science*, **275**, 1320-1323.

22. Gaggar, A., Shayakhmetov, D.M. and Lieber, A. (2003) CD46 is a cellular receptor for group B adenoviruses. *Nat Med*, **9**, 1408-1412.

23. Henaff, D., Salinas, S. and Kremer, E.J. (2011) An adenovirus traffic update: from receptor engagement to the nuclear pore. *Future microbiology*, **6**, 179-192.

24. Imperiale, M.J., Kao, H.T., Feldman, L.T., Nevins, J.R. and Strickland, S. (1984) Common control of the heat shock gene and early adenovirus genes: evidence for a cellular E1A-like activity. *Mol Cell Biol*, **4**, 867-874.

25. Lochmuller, H., Jani, A., Huard, J., Prescott, S., Simoneau, M., Massie, B., Karpati, G. and Acsadi, G. (1994) Emergence of early region 1-containing replication-competent adenovirus in stocks of replication-defective adenovirus recombinants (delta E1 + delta E3) during multiple passages in 293 cells. *Hum Gene Ther*, **5**, 1485-1491.

26. Chuah, M.K., Collen, D. and VandenDriessche, T. (2003) Biosafety of adenoviral vectors. *Curr Gene Ther*, **3**, 527-543.

27. Fallaux, F.J., Bout, A., van der Velde, I., van den Wollenberg, D.J., Hehir, K.M., Keegan, J., Auger, C., Cramer, S.J., van Ormondt, H., van der Eb, A.J. *et al.* (1998) New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Hum Gene Ther*, **9**, 1909-1917.

28. Schiedner, G., Hertel, S. and Kochanek, S. (2000) Efficient transformation of primary human amniocytes by E1 functions of Ad5: generation of new cell lines for adenoviral vector production. *Hum Gene Ther*, **11**, 2105-2116.

29. Campos, S.K. and Barry, M.A. (2007) Current advances and future challenges in Adenoviral vector biology and targeting. *Curr Gene Ther*, **7**, 189-204.

30. Havenga, M.J., Lemckert, A.A., Ophorst, O.J., van Meijer, M., Germeraad, W.T., Grimbergen, J., van Den Doel, M.A., Vogels, R., van Deutekom, J., Janson, A.A. *et al.* (2002) Exploiting the natural diversity in adenovirus tropism for therapy and prevention of disease. *J Virol*, **76**, 4612-4620.

31. Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol*, **26**, 808-816.

32. Holkers, M., Maggio, I., Liu, J., Janssen, J.M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T. and Gonçalves, M.A. (2013) Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*, **41**, e63.

33. Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X. *et al.* (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res*, **41**, 9049-9061.

34. Kim, H., Um, E., Cho, S.R., Jung, C., Kim, H. and Kim, J.S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods*, **8**, 941-943.

35. Pelascini, L.P., Maggio, I., Liu, J., Holkers, M., Cathomen, T. and Gonçalves, M.A. (2013) Histone deacetylase inhibition rescues gene knockout levels achieved with integrase-defective lentiviral vectors encoding zinc-finger nucleases. *Hum Gene Ther Methods*, **24**, 399-411.

36. Pelascini, L.P., Janssen, J.M. and Gonçalves, M.A. (2013) Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. *Hum Gene Ther*, **24**, 78-96.

37. He, T.C., Zhou, S., da Costa, L.T., Yu, J., Kinzler, K.W. and Vogelstein, B. (1998) A simplified system for generating recombinant adenoviruses. *Proc Natl Acad Sci U S A*, **95**, 2509-2514.

4

38. Janssen, J.M., Liu, J., Skokan, J., Gonçalves, M.A. and de Vries, A.A. (2013) Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. *J Gene Med*, **15**, 1-11.

39. Sambrook, J. and Russell, D.W. (2001) Molecular cloning: a laboratory manuel, 3rd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

40. Havenga, M.J., Holterman, L., Melis, I., Smits, S., Kaspers, J., Heemskerk, E., van der Vlugt, R., Koldijk, M., Schouten, G.J., Hateboer, G. *et al.* (2008) Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: high yield and preserved bioactivity. *Biotechnol Bioeng*, **100**, 273-283.

41. Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*, **36**, 59-74.

42. Fallaux, F.J., Kranenburg, O., Cramer, S.J., Houweling, A., Van Ormondt, H., Hoeben, R.C. and Van Der Eb, A.J. (1996) Characterization of 911: a new helper cell line for the titration and propagation of early region 1-deleted adenoviral vectors. *Hum Gene Ther*, **7**, 215-222.

43. Murakami, P. and McCaman, M.T. (1999) Quantitation of adenovirus DNA and virus particles with the PicoGreen fluorescent Dye. *Analytical biochemistry*, **274**, 283-288.

4

5

Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells

Ignazio Maggio*, Maarten Holkers*, Jin Liu, Josephine M. Janssen, Xiaoyu Chen & Manuel A.F.V. Gonçalves

Department of Molecular Cell Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands.

*These authors contributed equally to this work

ABSTRACT

CRISPR/Cas9-derived RNA-guided nucleases (RGNs) are DNA targeting systems, which are rapidly being harnessed for gene regulation and gene editing purposes in model organisms and cell lines. As *bona fine* gene delivery vehicles, viral vectors may be particularly fit to broaden the applicability of RGNs to other cell types including dividing and quiescent primary cells. Here, the suitability of adenoviral vectors (AdVs) for delivering RGN components into various cell types is investigated. We demonstrate that AdVs namely, second-generation fiber-modified AdVs encoding Cas9 or single guide RNA (gRNA) molecules addressing the Cas9 nuclease to the *AAVS1* "safe harbor" locus or to a recombinant model allele can be produced to high-titers (up to 20×10^{10} transducing units/ml). Importantly, AdV-mediated transduction of gRNA:Cas9 ribonucleoprotein complexes into transformed and non-transformed cells yields rates of targeted mutagenesis similar to or approaching those achieved by isogenic AdVs encoding TALENs targeting the same *AAVS1* chromosomal region. RGN-induced gene disruption frequencies ranged from 18% to 31% in a cell type-dependent manner. We conclude that AdVs constitute a valuable platform for introducing RGNs into human somatic cells regardless of their transformation status. This approach should aid investigating the potential and limitations of RGNs in numerous experimental settings.

Key words: CRISPR/Cas9; RNA-guided nucleases; Adenoviral vectors; Primary cells; Targeted gene knockout; Genome editing

INTRODUCTION

Genome engineering strategies based on designer nucleases can be harnessed to edit and to interrogate the function of genomic sequences in cells of higher eukaryotes. In common, these approaches entail the activation of cellular DNA repair pathways in target cells upon the induction of double-stranded DNA breaks (DSBs) at predefined chromosomal positions[1,2]. The engagement of non-homologous end-joining (NHEJ) and homologous recombination (HR) machineries at these artificially created DSBs can result in targeted gene knockout and addition, respectively. The potential of genome engineering is considerable for fundamental and applied research. For instance, in the form of new molecular medicine modalities such "genome surgery" approaches might constitute a watershed departure from canonical gene therapy in which chromosomal insertion of the therapeutic DNA is unpredictable[3].

Programmable nuclease technologies are developing at a rapid pace[1,2]. These include zinc-finger nucleases (ZFNs), engineered homing endonucleases, transcription activator-like effector nucleases (TALENs) and, more recently, RNA-guided nucleases (RGNs). RGNs are RNA-dependent DNA nucleases built on components from type II clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) systems. These systems have been evolving in bacteria and archaea as adaptive immune mechanisms against invading nucleic acids of foreign agents such as plasmids and phages[4,5]. CRISPR loci contain an array of short palindromic repeats separated by protospacer DNA snippets of around 20 bp that have been acquired from invading agents. Following expression, the short palindromic repeats in precursor CRISPR transcripts (pre-crRNA) hybridize to a portion of a trans-activating CRISPR RNA (tracrRNA). The bound tracrRNA, in concert with RNase III and Cas9, lead to the processing of the pre-crRNA into small CRISPR RNAs (crRNAs) each of which encompasses a sequence complementary to a specific protospacer. Finally, crRNAs base-paired to tracrRNAs loaded with Cas9 nucleases direct site-specific DNA cleavage following crRNA:target DNA hybridization. Of note, the fact that foreign DNA harbors a so-called protospacer adjacent motif (PAM) next to the protospacer sequence whereas crRNAs lack this motif, ensures that crRNA:tracrRNA:Cas9 complexes recognize exclusively the incoming exogenous DNA as opposed to the chromosomal CRISPR-incorporated protospacers[4,5].

The above-described fundamental insights combined with enticing work done *in vitro* and in bacteria[6,7] led the type II CRISPR/Cas9 system from *Streptococcus pyogenes* to be rapidly adapted as a gene editing toolbox for mammalian cells[8-10]. These studies comprised the transfection of cell lines with plasmids expressing from RNA polymerase II (Pol-II) and Pol-III promoters, respectively, Cas9 proteins and chimeric single guide RNAs (gRNAs) engineered by fusing sequence-tailored crRNAs to tracrRNA scaffolds. Thus, gRNAs are bifunctional molecules[7] consisting of crRNA targeting and tracrRNA scaffolding moieties. Owing to the fact that RGN target site specificity is governed by RNA-DNA hybridizations, as opposed

to protein-DNA interactions, make CRISPR/Cas-based systems unique among customizable nucleases. Importantly, this feature allows for facile nuclease retargeting and multiplexing by simply replacing the RNA component of each ribonucleoprotein complex[1,11]. However, independent studies carried out in cell lines indicate that RGNs display high levels of off-target DNA cleaving activities especially when compared to protein-based platforms such as TALENs with a tradeoff existing between activity and specificity[12-16]. The off-target cleaving profile of programmable nucleases is, clearly, a parameter to take into consideration in the context of both therapeutic and basic-research protocols.

To expand the range of experimental systems in which RGNs can be tested and optimized, it is crucial to introduce these rather large macromolecular complexes not only into established cell lines but also into non-transformed diploid cell types (i.e. Cas9: 1,368 aa; 158 kDa; *Cas9* ORF: 4.1 kb). In this regard, adenoviral vectors (AdVs) constitute interesting RGN delivery candidates especially owing to their episomal nature, large cloning capacity, high-titers and ability to transduced dividing as well as non-dividing cells[17].

In the present work, we sought to investigate the capacity of AdVs for packaging and delivering functional *gRNA* and human codon-optimized *Cas9* transcriptional units into human cells. To this end, we selected previously validated gRNA molecules whose sequences target Cas9 to the so-called *AAVS1* "safe harbor" locus located on the long arm of the human chromosome 19 (19q13.42-qter) or to a recombinant model allele consisting of a chromosomally integrated *eGFP* expression unit[10]. In parallel, to serve as a reference, we took along in these transduction experiments isogenic AdVs encoding TALENs, which have also been previously validated and target the same *AAVS1* region[18]. By deploying targeted mutagenesis assays as well as live-cell readouts based on fluorescence microscopy and flow cytometry, we demonstrate that AdVs can efficiently transfer functional RGN components into human cells regardless of their transformation status.

RESULTS

Adenoviral vectors encoding RGN components can be produced at high-titers
We started by investigating the feasibility of producing AdVs containing *Cas9* and *gRNA* transcriptional units. For this purpose, we selected *E1*- and *E2A*-deleted AdVs displaying fiber motifs recognizing the ubiquitous cellular receptor CD46 (Fig. 1A). This retargeted second-generation AdV system offers several advantages when compared to its conventional *E1*-deleted AdV counterpart including higher genetic capacity, lower levels of "leaky" viral gene expression and efficient transduction of cells lacking, or with low amounts of, the coxsackievirus-adenovirus receptor CAR on their surface. Among these are human cell types often utilized in experimental and applied research settings such as hematopoietic cells,

skeletal muscle progenitor cells (hereinafter referred to as myoblasts), and bone marrow-derived mesenchymal stromal cells (hMSCs)[19-22].

In the current work, we selected the previously validated gRNA_AAVS1-T2 unit[10], herein named gRNA[S1], to edit endogenous chromosomal sequences via AdV delivery of CRISPR/Cas9-derived RGNs. The nucleotide sequence of this gRNA targets the Cas9 nuclease to the human *AAVS1* locus (Fig. 1B). To serve as a reference, we took along the optimized designer nuclease pair TALEN-L[S1]/TALEN-R[S1] specific for the same *AAVS1* region[18] (Fig. 1B) and placed their ORFs under the control of the same transcriptional elements as those regulating *Cas9* expression. To this end, AdV shuttle plasmids expressing *Cas9*, *TALEN-L[S1]* or *TALEN-R[S1]* from the Pol-II *PGK* promoter and short *gRNA[S1]* molecules from the Pol-III *U6* promoter were constructed. The choice for the former promoter was motivated by its ubiquitous expression in a diverse array of human cell types and, owing to its human origin, higher resistance to silencing phenomena than that offered by commonly-used viral transcriptional elements.

Transient transfection experiments using as readout a surrogate DSB formation assay dependent on the NHEJ DNA repair pathway, showed that co-transfecting test cells with gRNA[S1]- and Cas9-encoding AdV shuttle plasmids resulted in *AAVS1* targeted mutagenesis (Fig. 2). The functionality of the AdV shuttle plasmids encoding the *AAVS1*-specific TALENs was also confirmed through similar transient transfection experiments (not shown). Next, we used these validated reagents to assemble full-length AdV molecular clones for rescuing and propagating the *E1*- and *E2A*-deleted (AdV$^{\Delta 2}$) fiber-modified (F^{50}) vectors AdV$^{\Delta 2}$P.Cas9.F^{50}, AdV$^{\Delta 2}$U6.gRNA^{S1}.F^{50}, AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50}. Following the rescue, propagation and purification of AdV particles, we determined that the AdV functional titers in the resulting preparations were within the typical range expected for these research-scale production rounds (i.e. 23.7×10^{10}, 13.3×10^{10}, 2.4×10^{10} and 4.2×10^{10} TU/ml for AdV$^{\Delta 2}$P.Cas9.F^{50}, AdV$^{\Delta 2}$U6.gRNA^{S1}.F^{50}, AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50}, respectively). In addition, as previously shown for AdVs harboring *TALEN* expression units[18], structural analysis of packaged AdV genomes isolated from purified particles of AdV$^{\Delta 2}$P.Cas9.F^{50} and AdV$^{\Delta 2}$U6.gRNA^{S1}.F^{50} revealed restriction patterns diagnostic for AdV backbones containing intact transgenes (Fig. 3).

5

Figure 1. Schematic representation of the genome-modifying AdVs deployed in this study and corresponding target sequences. (A) The AdVs encode the Cas9 nuclease or TALEN proteins specific for the human "safe harbor" locus *AAVS1* within *PPP1R12C* or gRNA molecules addressing Cas9 to *AAVS1* or to *eGFP* sequences. All vectors were assembled on the basis of a second-generation (i.e. *E1*- plus *E2A*-deleted) AdV backbone (AdV$^{\Delta 2}$) coding for chimeric fibers composed of basal domains from prototypic human adenovirus serotype 5 fused to the apical shaft and knob motifs from human adenovirus serotype 50 (F5/50). The *Cas9* and *TALEN* ORFs are under the transcriptional control of the human *PGK-1* gene promoter (PGK) and the simian virus 40 (SV40) polyadenylation signal whilst the *gRNAs* are regulated by a RNA Pol-III promoter (*U6*) and terminator sequences. Ψ, human adenovirus serotype 5 packaging signal; Ad L-ITR and Ad R-ITR, "left" and "right" adenoviral inverted terminal repeat, respectively. The structure of the single guide RNA molecules gRNA^{S1} and gRNA^{GFP}, composed of Cas9 targeting and scaffolding portions, are depicted base-paired to their respective *AAVS1* and *eGFP* target sequences. TALE-L and TALE-R, *AAVS1*-binding transcription activator-like effector repeats of designer nucleases TALEN-L and TALEN-R, respectively (TALE portions coding for repeat variable di-residues [RVDs] recognizing A, T, C and G are color-coded); FokI, nuclease domain of the type IIS restriction enzyme FokI; PAM, protospacer adjacent motif. **(B)** Location of the *AAVS1* target sites for the RNA- and TALE-guided nuclease complexes gRNA^{S1}:Cas9 and TALEN-L^{S1}:TALEN-R^{S1}, respectively, in relation to the whole *PPP1R12C* gene. The target sequences for the TALEN-L^{S1} and TALEN-R^{S1} proteins are in upper case, whereas that for the gRNA^{S1} is boxed (i.e. G[N]$_{19}$NGG). PAM, protospacer adjacent motif. Solid and open vertical arrowheads indicate the positions at which DNA nicking takes place catalyzed by the two nuclease domains of Cas9. ▶
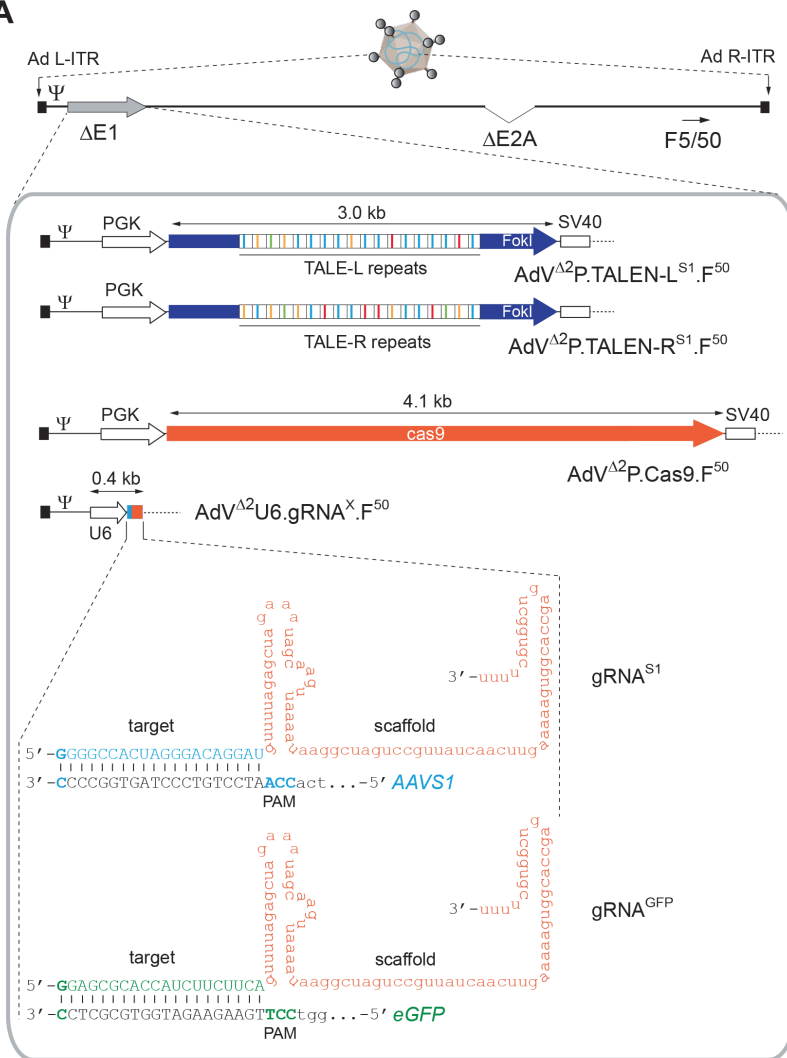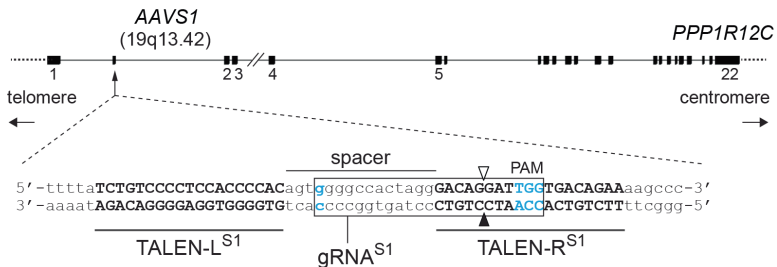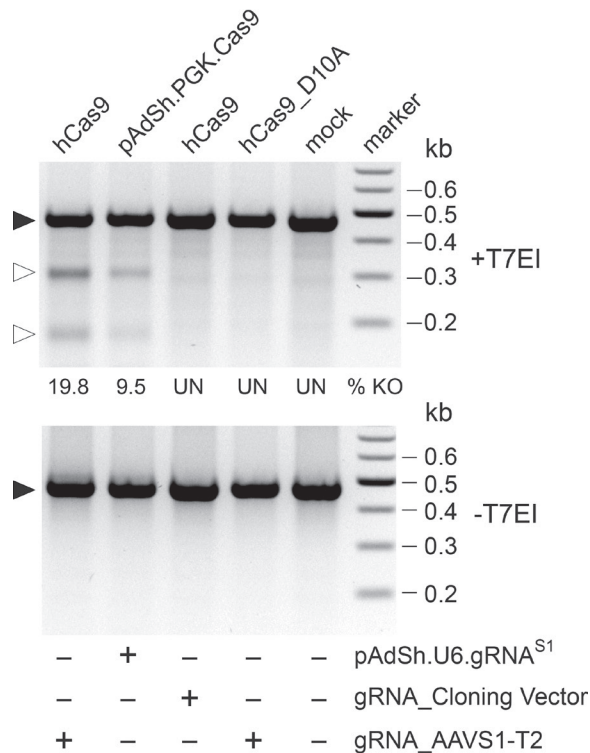
Figure 2. Transient transfection experiments for validating AdV shuttle plasmids pAdSh.PGK.Cas9 and pAdSh.U6.gRNA[S1]. *AAVS1*-specific PCR products amplified from *E1*-transformed PER.C6 cells co-transfected with hCas9 (9.6 kb) and gRNA_AAVS1-T2 (4 kb), pAdSh.PGK.Cas9 (11.6 kb) and pAdSh.U6.gRNA[S1] (7.3 kb), hCas9 and "empty gRNA" construct gRNA_Cloning Vector (3.9 kb), hCas9_D10A (9.6 kb) and gRNA_AAVS1-T2 or mock-transfected. Marker, Gene Ruler DNA Ladder (Fermentas). Plasmids hCas9 and hCas9_D10A express Cas9 nucleases which induce a DSB and a nick at *AAVS1*, respectively. After amplicon denaturation and reannealing, the presence of mismatches derived from NHEJ-mediated repair of site-specific DSBs *in cellula* was probed by T7 endonuclease I (T7EI) digestions (upper panel). Negative controls were provided by amplicons not exposed to T7EI (-T7EI) as well as T7EI-treated amplicons (+T7EI) corresponding to mock-transfected cells, to cells co-transfected with hCas9 and gRNA_Cloning Vector and to cells co-transfected with hCas9_D10A and gRNA_AAVS1-T2 encoding the "nickase" mutant version of Cas9 (i.e. Cas9[D10A]) and gRNA[S1], respectively. Solid and open arrowheads indicate the positions of, respectively, undigested and T7EI-digested DNA fragments whose sizes are consistent with DSB formation at the *AAVS1* target site. % KO and UN, knockout frequency and undetected, respectively.
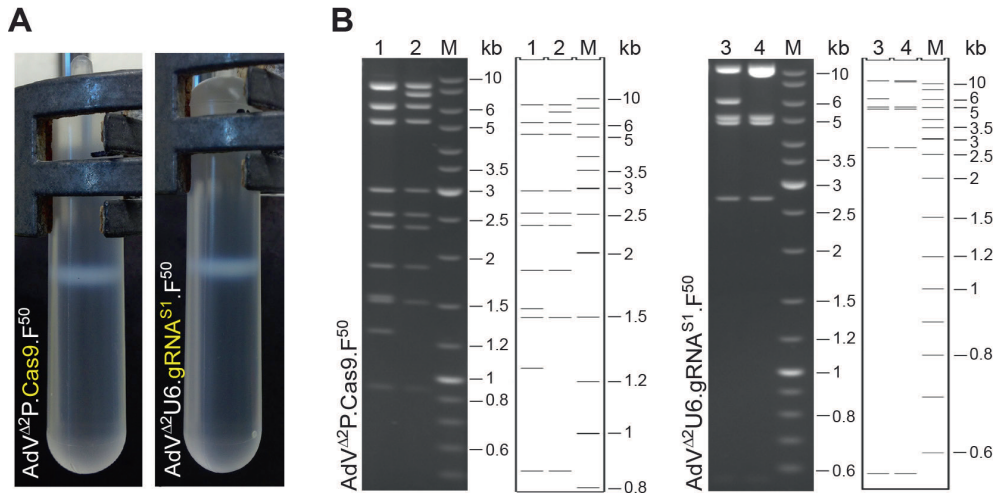
Figure 3. Structural analysis of AdV genomes. (A) Genome-containing AdV particles after ultracentrifugation banding through continuous CsCl density gradients (**B**) Characterization by restriction fragment length analysis of DNA isolated from CsCl-purified AdV particles. Left panels, in-gel and *in silico* restriction patterns of AdV$^{\Delta 2}$P.Cas9.F^{50} genomes (lanes 1) and respective pAdV$^{\Delta 2}$PGK.Cas9.F^{50} parental plasmid (lanes 2) treated with BglII. Right panel, in-gel and *in silico* restriction patterns of AdV$^{\Delta 2}$U6.gRNA^{S1}.F^{50} DNA (lanes 3) and corresponding pAdV$^{\Delta 2}$U6.sgRNA^{S1}.F^{50} parental plasmid (lanes 4) digested with Bsp1407I. Lanes M, Gene Ruler DNA Ladder marker.

AdVs deliver functional RGN complexes into human cells inducing robust site-specific mutagenesis at native and recombinant target chromosomal loci

Subsequently, we sought to test via transduction experiments followed by gene knockout assays, the functionality of the gRNA^{S1}- and Cas9-encoding AdVs in a panel of human cell types. These target cells comprised cervix carcinoma HeLa cells, osteosarcoma U2OS cells, hMSCs and myoblasts. In addition, we took advantage of the fine-tuned adenoviral infection mechanism to introduce into target cell nuclei a range of well-defined dosages of DNA templates encoding each RGN component. Related to this, it is noteworthy mentioning that, as of yet, bipartite CRISPR/Cas nuclease complexes have been introduced into target cells by transfection-based protocols mostly deploying fixed amounts of the two RGN components. In this regard it is relevant pointing out that in contrast to ZFNs and TALENs, in which the deployment of equimolar amounts of similarly sized monomers directly follows from their *modus operandi* (i.e assembly of catalytically active dimers at the target site), RGN activities might improve by optimizing total and relative quantities of *gRNA* and *Cas* templates. Therefore, for the experiments deploying AdV$^{\Delta 2}$P.Cas9.F^{50} and AdV$^{\Delta 2}$U6.gRNA^{S1}.F^{50} we set-up "co-transduction grids" to determine gene disruption levels resulting from different combined dosages as well as ratios of the two RGN elements. Transduction experiments in the HeLa and U2OS cell lines led to dose-dependent *AAVS1* disruption following AdV-mediated gene transfer as revealed by the detection of NHEJ-derived small insertions/deletions (indels) at

the target site (Fig. 4, Fig. 5A and Fig. 5B). Importantly, control transduction experiments in which only one of the two vectors was used did not result in detectable indel formation (Fig. 4). Moreover, co-transduction experiments in U2OS cells deploying side-by-side vector pairs AdV$^{\Delta2}$P.Cas9.F^{50}/AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50}/AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} showed that, in addition to a clear AdV dose dependency, the gene knockout frequencies achieved by gRNA^{S1}:Cas9 and TALEN-L^{S1}:TALEN-R^{S1} complexes were similar (compare Fig. 5B with the lower panel in Fig. 5C).
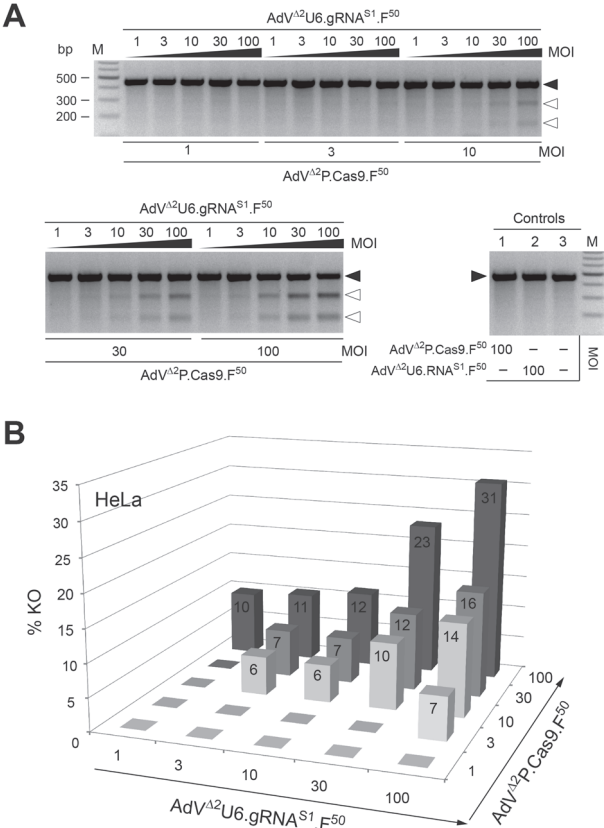
Figure 4. Transduction experiments in HeLa cells for the functional validation of AdV$^{\Delta2}$P.Cas9. F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50}. (A) Genotyping assays based on the detection of indels generated by NHEJ-mediated DSB repair at the chromosomal target site. The *AAVS1*-specific amplicons derived from genomic DNA of HeLa cells co-transduced with AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} at the indicated MOIs (in TU/cell) were denaturated and re-annealed, treated with T7EI and resolved through agarose gel electrophoresis. As negative controls, it was used genomic DNA isolated from HeLa cells exposed to 100 TU/cell of AdV$^{\Delta2}$P.Cas9.F^{50} (lane 1), 100 TU/cell of AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} (lane 2) or to no vector (lane 3). Solid and open arrowheads indicate the positions of undigested and T7EI-digested DNA molecules, respectively. Lane M, Gene Ruler DNA Ladder marker. **(B)** Target gene knockout frequencies in HeLa cells co-transduced with AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} as measured by densitometry of the DNA species depicted in Fig. 4A.
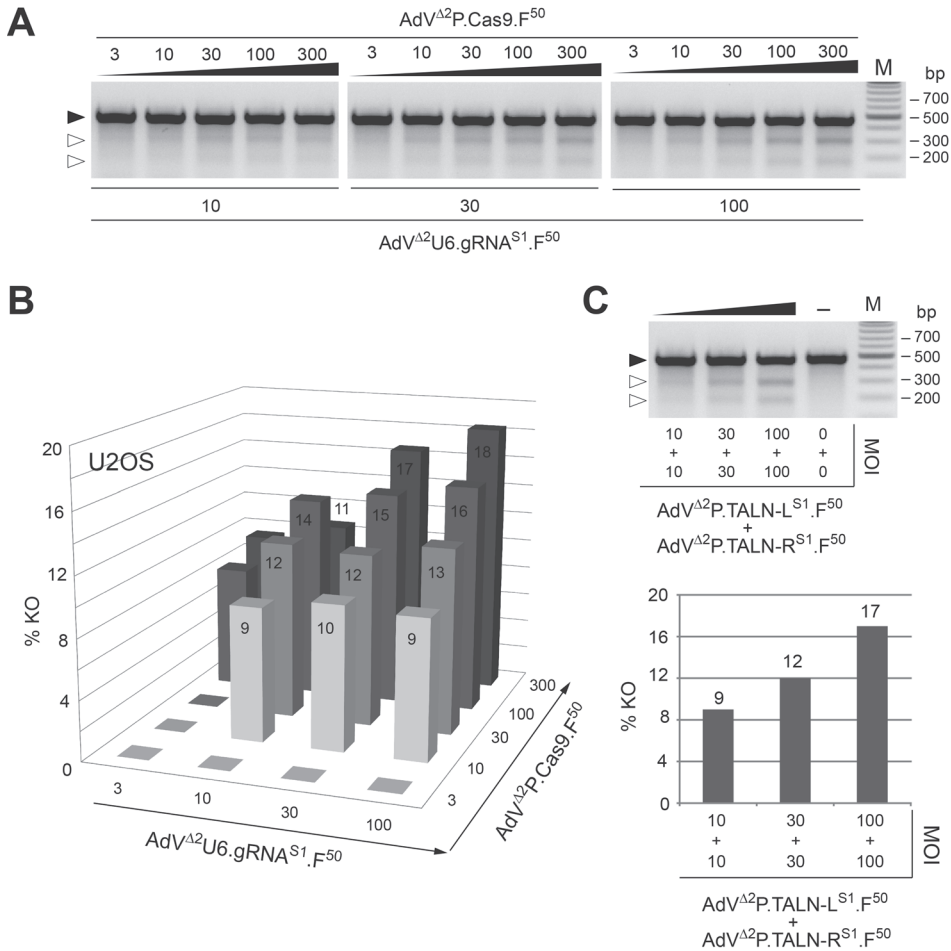
Figure 5. Analysis of site-specific DSB formation in U2OS cells following AdV-mediated delivery of transgenes encoding RGN and TALEN complexes. (**A**) Targeted mutagenesis experiments with AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50}. Agarose gel electrophoresis of *AAVS1*-specific PCR products amplified from chromosomal DNA of U2OS cells co-transduced with the indicated MOIs of AdV$^{\Delta2}$P. Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} (in TU/cell). Prior to loading the PCR products were denatured, re-annealed and treated with the mismatch-sensitive enzyme T7EI. Solid and open arrowheads point to the positions of undigested and T7EI-digested DNA molecules, respectively. Lane M, Gene Ruler DNA Ladder marker. (**B**) Target gene knockout frequencies in U2OS cells co-transduced with AdV$^{\Delta2}$P.Cas9. F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} as measured by densitometry of the DNA species shown in Fig. 5A. (**C**) Targeted mutagenesis experiments with AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}. Upper panel, T7EI-based assays on *AAVS1*-specific amplicons corresponding to genomic DNA of mock-transduced U2OS cells (-) or of U2OS cells co-transduced with the specified MOIs of AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} (in TU/cell). Solid and open arrowheads indicate the positions of undigested and T7EI-digested DNA molecules, respectively. Lanes M, Gene Ruler DNA Ladder marker. Lower panel, target gene knockout frequencies in U2OS cells co-transduced with AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P. TALEN-R^{S1}.F^{50} as determined by densitometry of the DNA fragments presented in the upper panel.

Follow-up transduction experiments in myoblasts generally yielded results similar to those obtained in HeLa and U2OS cells. Crucially, at matched MOIs, the levels of RGN-induced targeted mutagenesis were higher in myoblasts than in the two cancer-derived cell lines (Fig. 6A and Fig. 6B). For instance, 1:1 mixtures of AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} corresponding to a combined MOI of 60 TU/cell led to gene disruption frequencies in HeLa, U2OS and myoblasts of 12%, 12% and 27%, respectively (compare Fig. 4B and Fig. 5B with Fig. 6B). Importantly, myoblasts exposed to only one of the two RGN or TALEN pair members did not contain detectable indels at *AAVS1*. Finally, it is also noteworthy pointing out that at the high-end of the MOI scale tested in myoblasts (i.e. double dose of 30 TU/cell) the AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50}/AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} set yielded roughly a 2-fold higher gene knockout activity when compared to equivalent MOIs of the AdV$^{\Delta2}$P.Cas9.F^{50}/AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} pair (compare Fig. 6B with the lower panel in Fig. 6C). To establish the AdV transduction levels achieved in the CRISPR/Cas9-mediated targeted mutagenesis experiments corresponding to Figs. 4, 5 and 6, the various target cell types were exposed to particles of AdV$^{\Delta2}$U6. gRNA^{S1}.F^{50} and reporter vector AdV$^{\Delta2}$P.eGFP.F^{50} under the same conditions as those used in transduction experiments with AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} and AdV$^{\Delta2}$P.Cas9.F^{50}. Of note, the transgene regulatory elements in AdV$^{\Delta2}$P.eGFP.F^{50} are identical to those in AdV$^{\Delta2}$P.Cas9.F^{50}. Flow cytometric analyses clearly showed that the panel of MOIs selected for transducing each of the target cell types led to a wide range of gene delivery activities. Among these were MOIs yielding transduction levels of 100% (Supplementary Fig. 1).

In order to gather more information on target gene disruption activities attainable with the *AAVS1*-specific TALENs and RGN complexes within specific cell types and, simultaneously, probe for possible cell type-dependent effects on these activities following AdV-mediated gene transfer, we carried out parallel co-transduction experiments with both AdV sets on myoblast and hMSC cultures. In addition, to evaluate the impact of different relative concentrations of gRNA^{S1} and Cas9 molecules on gene knockout levels, we exposed these target cells to a constant total vector dose of 80 TU/cell spanning different ratios between AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50}.
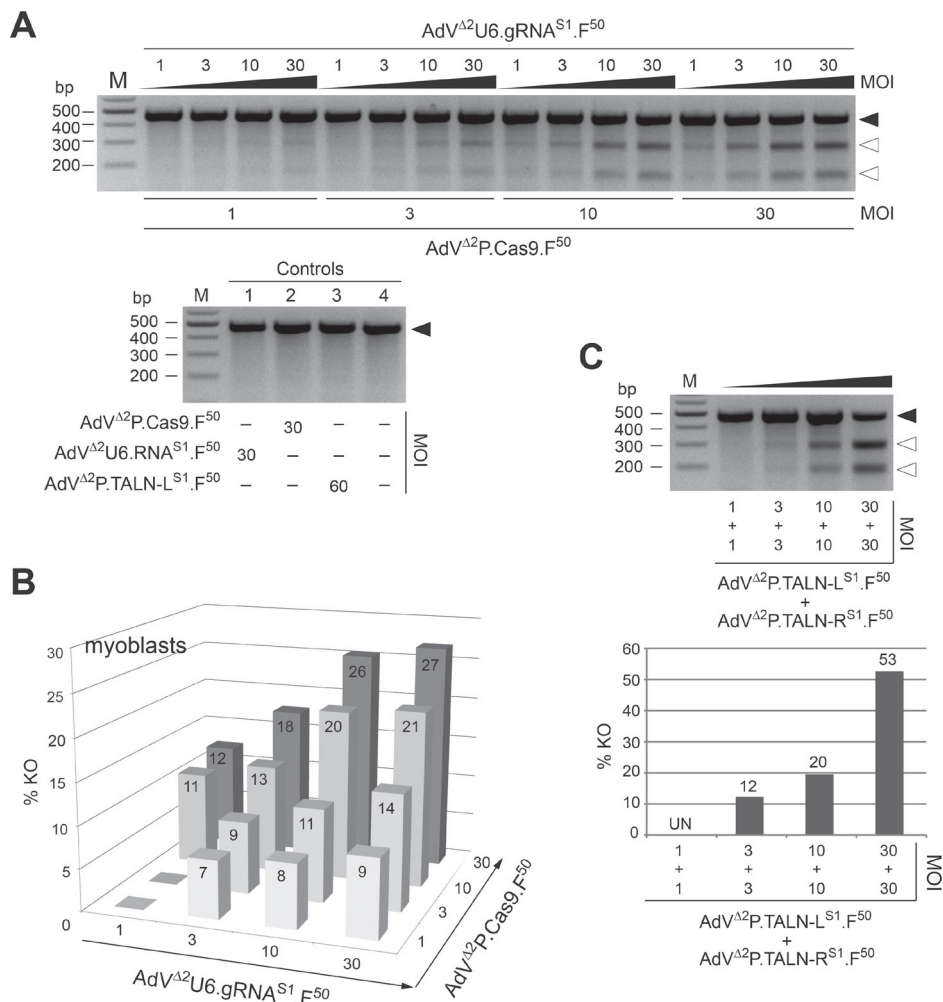
Figure 6. Analysis of targeted mutagenesis in human myoblasts following AdV-mediated transfer of DNA coding for RGN and TALEN complexes. (**A**) Targeted mutagenesis experiments with AdV$^{\Delta 2}$P.Cas9. F^{50} and AdV$^{\Delta 2}$.U6.gRNA^{S1}.F^{50}. Upper panel, *AAVS1*-specific PCR products corresponding to genomic DNA of myoblasts co-transduced with the indicated MOIs of AdV$^{\Delta 2}$P.Cas9.F^{50} and AdV$^{\Delta 2}$.U6.gRNA^{S1}. F^{50} (in TU/cell) subjected to T7EI-based assays. Lower panel, negative controls corresponding to human myoblasts that were singly transduced with AdV$^{\Delta 2}$P.Cas9.F^{50} (lane 1), AdV$^{\Delta 2}$.U6.gRNA^{S1}.F^{50} (lane 2) and AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} (lane 3) or that were mock-transduced (lane 4). (**B**) Target gene knockout frequencies in human myoblasts co-transduced with AdV$^{\Delta 2}$P.Cas9.F^{50} and AdV$^{\Delta 2}$.U6.gRNA^{S1}. F^{50} as measured by densitometry of the DNA species depicted in Fig. 6A. (**C**) Targeted mutagenesis experiments with AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50}. Upper panel, T7EI-based assays on *AAVS1*-specific amplicons obtained from genomic DNA of myoblasts co-transduced with the indicated MOIs of AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50} (in TU/cell). Solid and open arrowheads indicate the positions of undigested and T7EI-digested DNA molecules, respectively. Lanes M, Gene Ruler DNA Ladder marker. Lower panel, target gene disruption frequencies in myoblasts co-transduced with AdV$^{\Delta 2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta 2}$P.TALEN-R^{S1}.F^{50} as determined by densitometry of the DNA fragments shown in the upper panel.

The results depicted in Fig. 7A and in Fig. 7B corroborated the data on the higher TALEN pair activity in myoblasts presented in Fig. 6C and extended it to hMSCs. Indeed, gene knockout frequencies in myoblasts and hMSCs exposed to vector mixtures consisting of different ratios between AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} were consistently lower than those resulting from co-transductions with 1:1 mixtures of AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}. Of note, T7EI-digested DNA species corresponding to genomic DNA from cells transduced with the two nuclease systems displayed different migration profiles upon electrophoresis (Fig. 7A). These molecular weight differences are consistent with the predicted target site cleaving positions for the TALEN and RGN complexes i.e. within the spacer and three nucleotides "upstream" of the PAM, respectively (Fig. 1B). AdV transduction levels corresponding to these targeted mutagenesis experiments in myoblasts and hMSCs were determined by applying AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} and AdV$^{\Delta2}$P.eGFP.F^{50} (Fig. 7C) under the same experimental conditions as those used with AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} and AdV$^{\Delta2}$P.Cas9.F^{50} (Fig. 7B). Collectively, these experiments also demonstrated that, at equivalent MOIs, AdV-mediated delivery of the *AAVS1*-specific RGNs resulted in higher gene disruption levels in myoblasts than in hMSCs regardless of the proportions of each vector type in the co-transducing AdV$^{\Delta2}$P.Cas9.F^{50}/AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} mixtures (Fig. 7D).

To further test the AdV system as a delivery platform for CRISPR/Cas9-derived RGNs, AdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50} was generated. This vector has the genetic make-up of AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} except that it encodes the *eGFP*-specific gRNA^{GFP} in place of gRNA^{S1} (Fig. 1A). As with the other genome-modifying AdVs, AdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50} could also be produced to high titers of functional vector particles (i.e. 10.4×10^{10} TU/ml). Next, AdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50} was deployed together with AdV$^{\Delta2}$P.Cas9.F^{50} in transduction experiments in indicator H27 cells[23]. These cells contain a single-copy constitutionally active *eGFP* transcriptional unit providing, as a result, a convenient model system to evaluate target gene knockout strategies by live-cell tracing and quantification of gene-modified populations. Thus, as in experiments using AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50}, we set-up a "co-transduction grid" to determine targeted mutagenesis frequencies resulting from different combined dosages of AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50}. Direct fluorescence microscopy (Fig. 8A) as well as flow cytometry (Figs. 8B and 8C), revealed a clear dose-dependent increase in *eGFP* disruption levels varying from a minimum of 0.8% to a maximum of 65%. Controls consisting of H27 cells transduced either with AdV$^{\Delta2}$P.Cas9.F^{50} or with AdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50}, each applied at an MOI of 100 TU/cell, did not yield frequencies of eGFP-negative cells above background levels (Fig. 8B).

Taken our data together, we conclude that AdVs encoding RGN components can be produced to high titers and achieve robust targeted mutagenesis in a diverse array of human cell types.
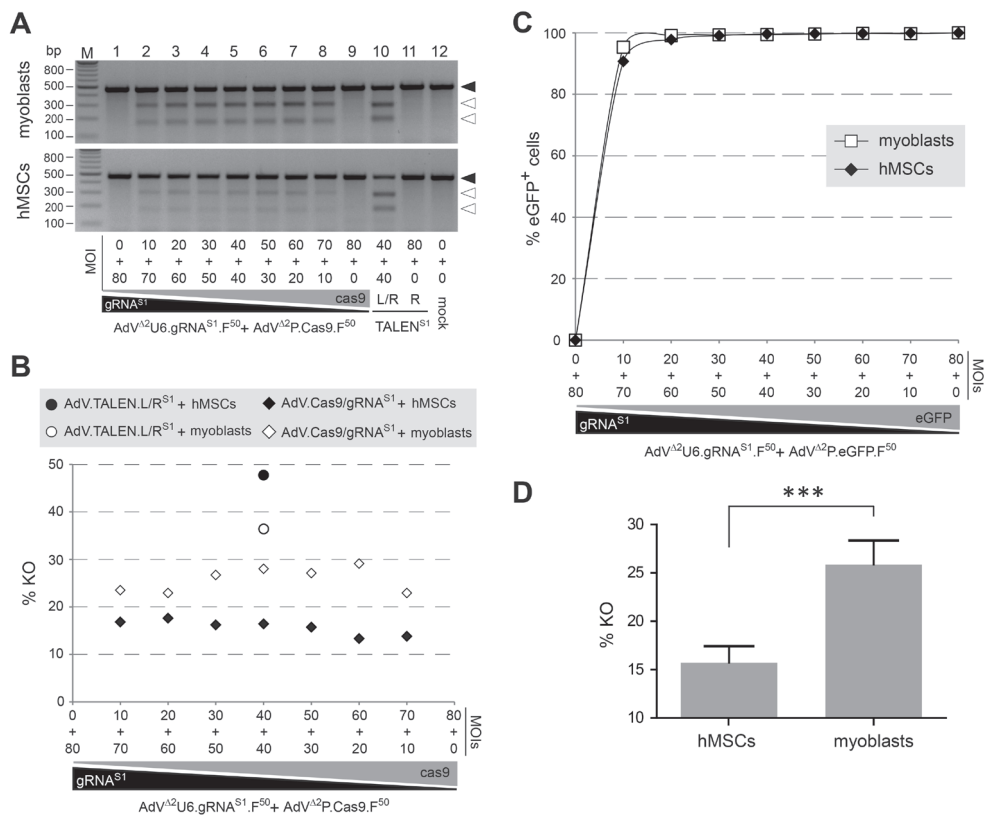
Figure 7. Analysis of target gene knockout frequencies following AdV-mediated introduction of gRNA^{S1}:Cas9 and TALEN-L^{S1}:TALEN-R^{S1} complexes into hMSCs versus human myoblasts. (A) Target site genotyping assays. Upper and lower panels, agarose gel electrophoresis of *AAVS1*-specific PCR products amplified from chromosomal DNA of myoblasts and hMSCs, respectively, co-transduced with the indicated MOIs of AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} (in TU/cell). Before loading the amplicons were denatured, re-annealed and treated with the mismatch-sensitive enzyme T7EI. The cells were co-transduced at a combined MOI of 80 TU/cell with different ratios of AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} or with 1:1 mixtures of AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}. Mock-transduced cells or cells exposed to 80 TU/cell of AdV$^{\Delta2}$P.Cas9.F^{50}, AdV$^{\Delta2}$.U6.gRNA^{S1}.F^{50} or AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50} served as negative controls. Solid and open arrowheads indicate the positions of undigested and T7EI-digested amplicons, respectively. Lanes M, Gene Ruler DNA Ladder marker. **(B)** Target gene knockout activities in hMSCs and myoblasts transduced with AdVs. Target gene disruption frequencies in cells co-transduced with AdV pairs AdV$^{\Delta2}$P.Cas9.F^{50}/AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} and AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50}/AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50} as determined by densitometry of DNA fragments depicted in Fig. 7A. **(C)** Gene delivery activities in hMSCs and myoblasts transduced with AdVs. Percentages of transduced cells incubated with different mixtures of reporter vector AdV$^{\Delta2}$P.eGFP.F^{50} and AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50}, applied at a combined MOI of 80 TU/cell. Cells exposed only to AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50} at an MOI of 80 TU/cell served as negative controls. The frequency of eGFP-positive cells was determined by flow cytometry at three days post-transduction. **(D)** Cumulative data corresponding to the gene knockout frequencies in hMSCs versus human myoblasts following AdV-mediated delivery of gRNA^{S1}:Cas9 complexes into these cells. Plotted data show the mean ± standard deviation (n=7).

Figure 8. Analyses of target gene editing frequencies in human H27 cells transduced with AdVs encoding *eGFP*-directed RGN complexes. (A) Fluorescence microscopy on cultures of *eGFP*-expressing H27 cells co-transduced with AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{GFP}.F^{50} applied at an MOI of **XX** TU/cell each. Negative controls consisted of mock-transduced H27 cells and of H27 cells transduced either with AdV$^{\Delta2}$P.Cas9.F^{50} or with AdV$^{\Delta2}$.U6.gRNA^{GFP}.F^{50} at an MOI of 100 TU/cell. Cell nuclei and *eGFP*-expressing cells in each microscopic field were identified via Hoechst 33342- and *eGFP*-directed fluorescence microcopy, respectively. Note the presence of eGFP-negative cells in cultures transduced with both vector types. **(B)** Flow cytometry-based quantification of target gene knockout levels in H27 cell cultures co-transduced with AdV$^{\Delta2}$P.Cas9.F^{50} and AdV$^{\Delta2}$.U6.gRNA^{GFP}.F^{50} at the indicated MOIs (in TU/cell). The fluorescence microscopy and flow cytometry were both performed at 15 days post-transduction.

DISCUSSION

One can postulate that AdV technologies, in addition to their utility in well-established biomedical research fields such as oncology and vaccinology[24-26], will start having an increasing role in experimental and applied genome engineering as well. An initial effort in this direction is provided by an AIDS therapeutic strategy based on knocking out the HIV-1 co-receptor gene *CCR5* following AdV-mediated *ZFN* transfer into CD4[+] T cells cultured *ex vivo*[27]. Related to this, we have recently shown that AdVs can also serve as delivery vehicles for introducing functional TALENs into human cells[18], whereas Ebina and colleagues have provided a proof-of-concept for using RGNs to disrupt HIV-1 proviral DNA[28]. In the current work, we have extended the utility of the AdV technology to genome engineering strategies based on RNA-dependent DNA nucleases derived from a type II CRISPR/Cas system.

Assuming that designer nuclease gene expression should ideally occur at high levels[29,30] and be short-term, AdVs may ultimately be preferable expression platforms when compared to other episomal viral vector systems such as silencing-prone integrase-defective lentiviral vectors (IDLVs)[30,31] and ssDNA adeno-associated viral vectors (AAVs)[32]. Moreover, the relatively large size of *Cas* ORFs may, to some extent, compromise the production of viral vectors whose optimal nucleic acid packaging capacities are significantly lower than those offered by AdV backbones and capsids. Still related to this payload capacity issue, our results further suggest that efficient chromosomal DSB formation seems to display a higher dependency on the overall RGN concentrations in target cell nuclei rather than on the use of a specific ratio between the Pol-II and Pol-III expression units. Thus, integrated vector designs in which both *Cas9* and *gRNA* expression units are co-delivered within single vector particles may turn out to be advantageous in certain experimental settings.

By carrying out transduction experiments with AdVs encoding validated gRNA[S1]:Cas9 complexes and optimized TALEN-L[S1]:TALEN-R[S1] heterodimers, we found that the latter sequence-specific nuclease system yielded gene disruption levels that were similar to or higher than those achieved by the former. This outcome is different from that reported recently by Ding and colleagues[33]. Through co-transfection experiments in human pluripotent stem cell lines comparing gRNA:Cas9 sets with TALEN pairs based on a specific architecture, these authors showed that, for each targeted locus, the RGNs consistently and substantially outperformed the TALENs. At this stage it is difficult to pinpoint any culprit(s) for these different outcomes since the variables are numerous and include the nature of the scaffolds for each platform, the delivery methods as well as the cell types and chromosomal target sequences tested. Nonetheless, our results do support the view that the TALEN platform is not *per se* inferior to that of RGNs in what site-specific chromosomal cleaving activity is concerned. Finally, we have shown that, at matched doses, AdV-mediated transfer of the *AAVS1*-specific RGNs led to higher gene disruption levels in myoblasts than in hMSCs.

5

Whether these data are a consequence of different epigenetic signatures at *AAVS1* or some other cell type-specific variable(s) remains to be investigated.

In conclusion, our findings demonstrate that AdVs constitute a valuable delivery option for introducing type II CRISPR/Cas-derived RGN complexes into human cells regardless of their transformation status and open as a result the perspective for RGN deployment and optimization in human primary cells. This research might include, in addition to on-target versus off-target assessments and gene knockout approaches, the testing of genome editing strategies based on donor DNA templates and AdV-mediated delivery of RGNs into target cell nuclei.

METHODS

Cells. The U2OS human osteosarcoma and the HeLa cervix carcinoma cells (American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% and 5% fetal bovine serum (FBS), respectively, at 37ºC in a 10% CO_2 atmosphere. The *E1*- and *E2A*-complementing AdV packaging cell line PER.E2A[34], was kept in DMEM supplemented with 10% FBS, 10 mM $MgCl_2$ and 250 μg/ml of Geneticin (Invitrogen). PER.E2A cells were regularly sub-cultured at 39ºC and were shifted to 34ºC during AdV production for the proper folding of their *E2A*-encoded thermo-sensitive DNA-binding protein. The origin of and culture conditions for the HeLa cell-derived clone H27, constitutively expressing *eGFP*, as well as those of myoblasts and hMSCs have been detailed elsewhere[23,35,36]. PER.C6 cells[37] were cultured at 37ºC in DMEM with 10% FBS and 10 mM $MgCl_2$.

Recombinant DNA. Standard recombinant DNA techniques were applied for the construction of the various AdV shuttle plasmids[38]. The TALEN-encoding plasmids pAdSh.PGK.TALEN-L[S1] [39] and pAdSh.PGK.TALEN-R[S1] [39] harbor the *TALEN* ORFs derived from 1383.pVAX.AAVS1. TALEN.L-94[18] and 1384.pVAX.AAVS1.TALEN.R-95[18], respectively. These constructs contain as transcription regulatory elements the human house-keeping phosphoglycerate kinase 1 gene (*PGK*) promoter and the simian virus 40 polyadenylation signal (SV40)[39]. The pAdSh. PGK.Cas9 and pSh.AAVS1.eGFP[40] constructs have these *PGK* and SV40 regulatory elements controlling the human codon-optimized ORF encoding hCas9[10] (herein referred to as Cas9) and the reporter eGFP, respectively. The *Cas9* ORF was isolated from Addgene plasmid 41815. The expression units gRNA_AAVS1-T2[10] and gRNA_GFP-T2[10] based on the *U6* RNA Pol III promoter (herein dubbed gRNA[S1] and gRNA[GFP], respectively) were retrieved from Addgene plasmid 41818 and 41820 and were inserted into the MCS of pAdSh.MCS.SV40, resulting in constructs pAdSh.U6.gRNA[S1] and pAdSh.U6.gRNA[GFP], respectively.

The *E1*- and *E2A*-deleted (i.e. second-generation) fiber-modified AdV molecular clones pAdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50}, pAdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}, pAdV$^{\Delta2}$P.Cas9.F^{50}, pAdV$^{\Delta2}$U6.gRNA^{S1}.F^{50}, pAdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50} and pAdV$^{\Delta2}$P.eGFP.F^{50} were assembled via homologous recombination after transformation of BJ5183$^{pAdEasy-2.50}$ *E. coli* cells[41] with MssI-treated AdV shuttle plasmids pAdSh.PGK.TALEN-L^{S1}, pAdSh.PGK.TALEN-R^{S1}, pAdSh.PGK.Cas9, pAdSh.U6.gRNA^{S1}, pAdSh.U6.gRNA^{GFP} and pSh.AAVS1.eGFP[40], respectively.

Validation of Cas9- and gRNA^{S1}-encoding AdV shuttle plasmids. PER.C6 cells were plated at a density of 1.5×10^6 cells per well of 6-well plates (Greiner Bio-One). The next day, a total amount of 6 µg of DNA corresponding to 1:1 mixtures of hCas9 (Addgene plasmid 41815) and gRNA_AAVS1-T2 (Addgene plasmid 41818), pAdSh.PGK.Cas9 and pAdSh.U6.gRNA^{S1}, hCas9 and gRNA_Cloning Vector (Addgene plasmid 41824) or hCas9_D10A (Addgene plasmid 41816) and gRNA_AAVS1-T2, were transfected by deploying a 1 mg/ml polyethyleneimine (PEI) solution (Polysciences) essentially as described elsewhere[41] except for the use of 6 µg of DNA and 19.7 µl of PEI instead of 6.25 µg of DNA and 18.8 µl of PEI. At 3 days post-transfection genomic DNA from mock-transfected cells and from co-transfected cells, was isolated according to a previously described method[42]. Targeted gene disruption was assessed by using the T7 endonuclease I (T7EI)-based assay as described below.

Viral vector production, purification and titration. The production of viral vectors AdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50}, AdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}, AdV$^{\Delta2}$P.Cas9.F^{50}, AdV$^{\Delta2}$U6.gRNA^{S1}.F^{50}, AdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50} and AdV$^{\Delta2}$P.eGFP.F^{50}, was carried out essentially as described elsewhere[41]. In brief, the rescue of AdV particles was initiated by transfecting 1.25×10^6 PER.E2A cells seeded in medium without Geneticin in wells of 6-well plates (Greiner Bio-One). The next day, the cells were transfected by using 18.8 µl of a 1 mg/ml PEI solution and 6.25 µg of PacI-linearized plasmids pAdV$^{\Delta2}$P.TALEN-L^{S1}.F^{50}, pAdV$^{\Delta2}$P.TALEN-R^{S1}.F^{50}, pAdV$^{\Delta2}$P.Cas9.F^{50}, pAdV$^{\Delta2}$U6.gRNA^{S1}.F^{50}, pAdV$^{\Delta2}$U6.gRNA^{GFP}.F^{50} and pAdV$^{\Delta2}$P.eGFP.F^{50}. Following an overnight incubation period at 39ºC, the transfection media was replaced and the cells were transferred to the permissive temperature of 34ºC. After the emergence of complete cytophatic effect (CPE), the producer cells were harvested and subjected to three cycles of freezing and thawing in liquid N$_2$ and a 37ºC water bath, respectively. Rescued AdVs present in clarified producer cell supernatants, were subsequently amplified through propagation on newly seeded PER.E2A cells. Large-scale AdV stocks were prepared by infecting producer cells at 70-80% confluence in 16 175-cm^2 cell cultures flasks (Greiner Bio-One). The details on the isopycnic CsCl density-gradient ultracentrifugation and ultrafiltration used to purify and de-salt AdV preparations, respectively, have been described before[41]. The titers of purified AdV stocks were determined by TCID$_{50}$ assays as follows. One hundred microliters of a PER.E2A cell suspension containing 5.0×10^4 cells/ml in DMEM containing 2% FBS and 10 mM MgCl$_2$, were dispensed in wells of

5

96-well plates (Greiner Bio-One). After an overnight incubation period at 39ºC in a 10% CO_2 atmosphere, the cells were exposed in octuplicate to 100 µl of 10-fold dilutions (range: 10^{-5} through 10^{-12}) of the various viral vector preparations. Next, the cells were incubated for 2.5 weeks at 34ºC in a 10% CO_2 atmosphere having 100 µl of DMEM containing 2% FBS and 10 mM $MgCl_2$ being added at 1 week post-infection. After this incubation period, wells with cultures containing viral plaques or at full CPE were scored to calculate the titers in terms of transducing units per ml (TU/ml) as specified elsewhere[41].

Isolation of viral vector DNA for structural analysis. The isolation of AdV genomes from purified particles was carried out as specified elsewhere[41]. The *in silico* restriction patterns were made with the aid of Gene Construction kit (version 2.5) software (Textco BioSoftware, Inc.).

Transduction experiments. Myoblasts, hMSCs and U2OS cells were seeded in their respective media at a density of 1×10^5 cells per well of 24-well plates (Greiner Bio-One), whereas HeLa cells were plated in the same vessel format at a density of 5×10^4 cells. Approximately 24 hours after seeding, the cells were exposed in 500-µl volumes to the two different AdV pair combinations (i.e. $AdV^{\Delta 2}P.TALEN-L^{S1}.F^{50}/AdV^{\Delta 2}P.TALEN-R^{S1}.F^{50}$ or $AdV^{\Delta 2}P.Cas9.F^{50}/AdV^{\Delta 2}U6.gRNA^{S1}.F^{50}$) at the indicated multiplicities of infection (MOI). Mock-transduced cells or, whenever indicated, cells transduced exclusively with one element of each vector pair, served as negative controls. Three days post-infection, total cellular DNA was isolated by using the DNeasy Blood & Tissue kit (Qiagen) following the manufacturer's recommendations.

Transduction experiments on indicator H27 cells. One day before transduction, 1.0×10^5 H27 cells were seeded in wells of 24-well plates. After the removal of the culture medium, the H27 cell monolayers were incubated in 500-µl volumes consisting of various total and relative amounts of $AdV^{\Delta 2}P.Cas9.F^{50}$ and $AdV^{\Delta 2}U6.gRNA^{GFP}.F^{50}$. Mock-transduced H27 cells and H27 cells exposed exclusively to $AdV^{\Delta 2}P.Cas9.F^{50}$ or to $AdV^{\Delta 2}U6.gRNA^{GFP}.F^{50}$ each applied at an MOI of 100 TU/cell, provided for negative controls. Subsequently, after extensive sub-culturing to remove reporter protein from cells with disrupted *eGFP* ORFs, the presence of eGFP-negative cells was assessed by direct fluorescence microscopy and flow cytometry at 15 days post-transduction.

Direct fluorescence microscopy. Targeted *eGFP* knockout in H27 cell cultures was monitored by direct fluorescence microscopy. The H27 cell nuclei were stained by adding Hoechst 33342 (Molecular Probes) at 10 µg/ml for 10 min. Next, the cell cultures were washed twice with PBS, after which cultured medium was added. The eGFP- and Hoechst 33342-specific

signals were detected by using an Olympus IX51 inverse fluorescence microscope (Olympus). Micrographs were captured and archived by deploying a XC30 Peltier CCD camera and the CellF software (both from Olympus).

Flow cytometry. The frequencies of eGFP-positive cells in cultures transduced with AdV$^{\Delta 2}$P. eGFP.F[50] and of eGFP-negative cells in H27 cell cultures were determined by using a BD LSR II flow cytometer (BD Biosciences). Data were analyzed with the aid of FlowJo 7.2.2 software (Tree Star). Mock-transduced H27 cells served to establish the background fluorescence. Samples of at least 10,000 viable single cells were analyzed per experimental condition.

T7 endonuclease I-based genotyping assays. Target gene disruption levels were assessed by using mismatch-sensitive T7EI endonuclease I. To this end, genomic DNA samples from target cells were subjected to PCR for the amplification of a 469-bp DNA segment encompassing the target sites for the gRNA[S1]:Cas9 and TALEN-L[S1]:TALEN-R[S1] complexes. The primers, PCR reagents and cycling conditions used have been described in detail before[18]. Next, the resulting amplicons were denatured and reannealed in a thermocycler by applying the program specified in Supplementary Table S1. One fifth of each PCR mixture was incubated in 15-μl reactions with 1× NEBuffer 2 and 5 U of T7EI (both from New England Biolabs). Control reactions lacked T7EI. After 15 minutes at 37ºC, the samples were subjected to electrophoresis through 2% (w/v) agarose in 1×Tris-Acetate-EDTA buffer. Finally, ethidium bromide-stained DNA species were measured in a Molecular Imager Gel-Doc™ XR+ with the aid of Bio-Rad Image Lab 4.1 software (both from Bio-Rad).

5

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

I.M. and M.H. contributed equally to this work. I.M. and M.H. performed the experiments. J.L., J.M.J. and X.C. produced and characterized reagents. I.M., M.H., J.M.J. and M.A.F.V.G. designed the experiments and analyzed the data. M.A.F.V.G. conceived and initiated the research. M.A.F.V.G. wrote the manuscript with the help from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Gaj, T., Gersbach, C.A. & Barbas, C.F.3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. **31**, 397-405 (2013).

2. Segal, D.J. & Meckler, J.F. Genome engineering at the dawn of the golden age. *Annu*. *Rev. Genomics Hum. Genet*. **14**, 135-158 (2013).

3. Biasco, L., Baricordi, C. & Aiuti, A. Retroviral integrations in gene therapy trials. *Mol. Ther*. **20**, 709-716 (2012).

4. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet*. **45**, 273-297 (2011).

5. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-338 (2012).

6. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* **109**, E2579-E2586 (2012).

7. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. & Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

8. Cong, L. *et al*. Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823 (2013).

9. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. & Doudna, J. RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).

10. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. & Church, G.M. RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

11. Gasiunas, G. & Siksnys, V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol*. **21**, 562-567 (2013).

12. Cradick, T.J., Fine, E.J., Antico, C.J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*. **41**, 9584-9592 (2013).

13. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. & Sander, J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol*. **31**, 822-826 (2013).

14. Hsu, P.D. *et al*. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. **31**, 827-832 (2013).

15. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. & Church, G.M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol*. **31**, 833-838 (2013).

16. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. & Liu, D.R. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*. **31**, 839-843 (2013).

17. Gonçalves, M.A.F.V & de Vries, A.A.F. Adenovirus: from foe to friend. *Rev Med. Virol*. **16**, 167-186 (2006).

18. Holkers, M., Maggio, I., Liu, J., Janssen, J.M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T. & Gonçalves, M.A.F.V. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. **41**, e63 (2013).

19. Shayakhmetov, D.M., Papayannopoulou, T., Stamatoyannopoulos, G. & Lieber, A. Efficient gene transfer into human CD34(+) cells by a retargeted adenovirus vector. *J. Virol*. **74**, 2567-2583 (2000).

5

20. Knaän-Shanzer, S., van der Velde, I., Havenga, M.J., Lemckert, A.A., de Vries, A.A.F. & Valerio, D. Highly efficient targeted transduction of undifferentiated human hematopoietic cells by adenoviral vectors displaying fiber knobs of subgroup B. *Hum. Gene Ther*. 12, 1989-2005 (2001). Erratum in: *Hum. Gene Ther*. **14**, 1214 (2003).

21. Knaän-Shanzer, S., van de Watering, M.J., van der Velde, I., Gonçalves, M.A.F.V., Valerio, D. & de Vries, A.A.F. Endowing human adenovirus serotype 5 vectors with fiber domains of species B greatly enhances gene transfer into human mesenchymal stem cells. *Stem Cells* **23**, 1598-1607 (2005).

22. Gonçalves, M.A.F.V., Holkers, M., Cudré-Mauroux, C., van Nierop, G.P., Knaän-Shanzer, S., van der Velde, I., Valerio, D. & de Vries, A.A.F. Transduction of myogenic cells by retargeted dual high-capacity hybrid viral vectors: robust dystrophin synthesis in duchenne muscular dystrophy muscle cells. *Mol. Ther*. **13**, 976-986 (2006).

23. Gonçalves, M.A.F.V., van der Velde, I., Knaän-Shanzer, S., Valerio, D. & de Vries, A.A.F. Stable transduction of large DNA by high-capacity adeno-associated virus/adenovirus hybrid vectors. *Virology* **321**, 287-296 (2004).

24. Barouch, D.H. & Nabel, G.J. Adenovirus vector-based vaccines for human immunodeficiency virus type 1. *Hum. Gene Ther*. **16**, 149-156 (2005).

25. Post, D.E., Khuri, F.R., Simons, J.W. & van Meir, E.G. Replicative oncolytic adenoviruses in multimodal cancer regimens. *Hum. Gene Ther*. **14**, 933-946 (2003).

26. Aurisicchio, L. & Ciliberto, G. Genetic cancer vaccines: current status and perspectives. *Expert. Opin. Biol. Ther*. **12**, 1043-1058 (2012).

27. Perez, E.E. *et al*. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol*. **26**, 808-816 (2008).

28. Ebina, H., Misawa, N., Kanemura, Y. & Koyanagi Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep*. **3**, 2510 (2013).

29. Kim, H., Um, E., Cho, S.R., Jung, C., Kim, H. & Kim, J.S. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods* **8**, 941-943 (2011).

30. Pelascini, L.P., Maggio, I., Liu, J., Holkers, M., Cathomen, T. & Gonçalves, M.A.F.V. Histone Deacetylase Inhibition Rescues Gene Knockout Levels Achieved with Integrase-Defective Lentiviral Vectors Encoding Zinc-Finger Nucleases. *Hum. Gene Ther. Methods* **24**, 399-411 (2013).

31. Pelascini, L.P., Janssen, J.M., & Gonçalves, M.A.F.V. Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. *Hum. Gene Ther*. **24**, 78-96 (2013).

32. Ferrari, F.K., Samulski, T., Shenk, T. & Samulski, R.J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol*. **70**, 3227-3234 (1996).

33. Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A. & Musunuru, K. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393-394 (2013).

34. Havenga, M.J. *et al*. Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: high yield and preserved bioactivity. *Biotechnol. Bioeng*. **100**, 273-283 (2008).

35. Cudré-Mauroux, C., Occhiodoro, T., König, S., Salmon, P., Bernheim, L. & Trono, D. Lentivector-mediated transfer of Bmi-1 and telomerase in muscle satellite cells yields a duchenne myoblast cell line with long-term genotypic and phenotypic stability. *Hum. Gene Ther*. **14**, 1525-1533 (2003).

36. Gonçalves, M.A.F.V., Janssen, J.M., Nguyen, Q.G., Athanasopoulos, T., Hauschka, S.D., Dickson, G. & de Vries, A.A.F. Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. *Mol. Ther*. **19**, 1331-1341 (2011).

37. Fallaux, F.J. *et al*. New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Hum. Gene Ther*. **9**, 1909-1917 (1998).

38. Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Manual 3rd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001).

39. Holkers, M., Cathomen, T. & Gonçalves, M.A.F.V. Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells. *Methods* S1046-2023 (2014). Epub ahead of print.

40. Coluccio, A. *et al*. Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther*. **21**, 1695-1704 (2013).

41. Janssen, J.M., Liu, J., Skokan, J., Gonçalves, M.A.F.V. & de Vries, A.A.F. Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. *J. Gene Med*. **15**, 1-11 (2013).

42. Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R. & Berns, A. Simplified mammalian DNA isolation procedure. *Nucleic Acids Res*. **19**, 4293 (1991).

5

Adenoviral vector DNA for accurate genome editing with engineered nucleases

Maarten Holkers[1,5], Ignazio Maggio[1,5], Sara F.D. Henriques[1,2], Josephine M. Janssen[1],
Toni Cathomen[3,4] & Manuel A.F.V. Gonçalves[1]

[1]Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, the
Netherlands. [2]Departamento de Biologia Animal, Faculdade de Ciências da Universidade
de Lisboa, Lisbon, Portugal. [3]Institute for Cell and Gene Therapy, University Medical Center
Freiburg, Freiburg, Germany. [4]Center for Chronic Immunodeficiency, University Medical
Center Freiburg, Freiburg, Germany. [5]These authors contributed equally to this work.
Correspondence should be addressed to M.A.F.V.G. (m.goncalves@lumc.nl).

ABSTRACT

Engineered sequence-specific nucleases and donor DNA templates can be customized to edit mammalian genomes via the homologous recombination (HR) pathway. Here we report that the nature of the donor DNA greatly affects the specificity and accuracy of the editing process following site-specific genomic cleavage by transcription activator–like effector nucleases (TALENs) and clustered, regularly interspaced, short palindromic repeats (CRISPR)-Cas9 nucleases. By applying these designer nucleases together with donor DNA delivered as protein-capped adenoviral vector (AdV), free-ended integrase-defective lentiviral vector or nonviral vector templates, we found that the vast majority of AdV-modified human cells underwent scarless homology-directed genome editing. In contrast, a significant proportion of cells exposed to free-ended or to covalently closed HR substrates were subjected to random and illegitimate recombination events. These findings are particularly relevant for genome engineering approaches aiming at high-fidelity genetic modification of human cells.

INTRODUCTION

The exchange of genetic information between native acceptor loci and exogenous donor DNA through error-free HR is an established strategy to manipulate prokaryote and eukaryote genomes with nucleotide-level precision[1]. However, in mammalian somatic cells, typical frequencies of spontaneous HR-mediated gene-targeting range from 10^{-8} to 10^{-6} events per transfected cell, with most exogenous DNA being found randomly integrated throughout host-cell chromosomes[2–4]. Importantly, the deployment of sequence-specific nucleases greatly increases the odds of retrieving cells with specific allelic alterations[5]. Generation of double-stranded DNA breaks (DSBs) at predefined chromosomal positions together with the introduction of donor DNA containing sequences identical to those bracketing the genomic lesion can increase gene targeting by several orders of magnitude. In gene therapy, for instance, inserting transcriptional units into specific genomic positions (i.e., so-called safe harbors) or directly repairing faulty genes within their native chromosomal context is a highly desirable goal[6]

Viral vectors constitute attractive gene delivery vehicles owing to their efficient transduction of a wide range of cell types. In addition, in contrast to bulk nucleic acid transfections[7], viral transductions permit fine control over the number of DNA copies that reach the nucleus. We have investigated the role of the donor-delivering vectors on the specificity and accuracy of the DNA editing process. Specificity refers to the relative frequencies of on-target versus off-target insertions; accuracy refers to the structure or arrangement of site-specifically integrated exogenous DNA. Low-fidelity gene targeting includes the chromosomal integration of exogenous DNA copies in tandem (i.e., concatemers) as well as the incorporation of virus-derived sequences at the target site. These events may result from the generation of vector-vector or vector-host DNA junctions through error-prone nonhomologous end joining. Clearly, accurate site-specific gene addition and repair should yield, respectively, homogenous transgene activities and restored open reading frames (ORFs) in target cell populations.

In this study, we compared the specificity and the accuracy of nuclease-induced gene targeting upon the delivery of various types of HR substrates (i.e., protein-capped, free-ended and covalently closed DNA) and found that the nature of the donor DNA greatly affects these key parameters. We demonstrate that donor DNA transferred by protein-capped AdVs is amenable to homology-directed gene targeting after sequence-specific genomic cleavage by TALENs and CRISPR-Cas9 nucleases. Further, donor delivery by AdVs results in diminished off-target chromosomal insertion, concatemeric 'footprint' formation and prokaryotic DNA incorporation. In consequence, genetically modified cell populations generated via AdV donor DNA transfer show homogenous transgene activity.

RESULTS

Designer nuclease–induced IDLV DNA targeting is inaccurate

Integrase-defective lentiviral vectors (IDLVs)[8] are one of the most commonly used viral vectors for the delivery of HR substrates into human cells[9–14]. Because TALENs display a particularly favorable specificity profile[15], we started by investigating the specificity and fidelity of TALEN-induced chromosomal insertion of IDLV donor DNA. We introduced TALENs specific for the safe harbor *AAVS1* locus into human myoblasts through the early region 1 (*E1*)-deleted adenoviral vectors AdV.TALEN-L[S1] and AdV.TALEN-R[S1] (ref. 16) together with the target site-matched HR substrate delivered by IDLV.donor[S1] particles (**Fig. 1a**). The AdV-delivered TALEN pair TALEN-L[S1] and TALEN-R[S1] forms a dimeric nuclease complex by binding, respectively, at the 'left' (L) and 'right' (R) end of the AAVS1 target sequence (S1). The IDLV.donor[S1], generated on the basis of plasmid AQ25_pLV.donor[S1] (Supplementary Note), contains a transgene encoding the fluorescent reporter EGFP flanked by sequences sharing identity to DNA bracketing the TALEN target site. Cells exposed to the latter vector alone or mixed with AdV.TALEN-L[S1] served as negative controls. After subculturing to eliminate episomal vector DNA, flow cytometry (**Fig. 1b**) and live-cell fluorescence microscopy (**Supplementary Fig. 1a**) showed a significant (*P* < 0.0001) nuclease-dependent increase in the frequencies of stably transduced cells.

Next, to gauge the relative frequencies of on-target versus off-target IDLV donor DNA integration, we randomly selected EGFP[+] myoblast clones (*n* = 104) from cultures exposed to the *AAVS1*-specific TALENs. We observed, using PCR screening with primers designed to yield amplicons diagnostic for HR-derived junctions between foreign and native target DNA (**Fig. 1a** and **Supplementary Fig. 1b**), that 86.6% of the EGFP[+] cells underwent homology-directed chromosomal integration of the exogenous DNA (**Fig. 1c** and **Supplementary Fig. 1b**). The resulting *AAVS1*-donor DNA junctions represented events involving the telomeric (6.7%), centromeric (3.9%) or both ends of the targeting template (76.0%) (**Fig. 1c**). Further characterization of IDLV integrants revealed high frequencies of head-to-tail (H-T) concatemeric forms not only in the nontargeted but also in the three *AAVS1*-targeted clonal fractions (38.5%) (**Fig. 1c** and **Supplementary Fig. 1b**). Of note, PCR analysis of EGFP[+] cells sorted from cultures that were not exposed to TALENs did not yield amplicons diagnostic for homology-directed gene addition (**Supplementary Fig. 1c**). These data indicate that a substantial fraction of incoming IDLV donor templates integrates randomly into host cell chromosomes, presumably at spontaneous DSBs through a noncanonical, i.e., integrase-independent, process[17,18]. Moreover, owing to their concatemeric structure, a sizable proportion of IDLV-exposed cells harbors unwanted HIV-derived *cis*-acting elements (**Fig. 1c** and **Supplementary Figs. 1b** and **2**). These tandem repeats are expected to neither restore endogenous ORFs nor yield homogeneous transgene expression levels in the context of

gene repair and gene addition strategies, respectively. These results are in line with those of other experiments carried out by us (**Supplementary Results** and **Supplementary Fig. 3**) as well as by others[11,12,14] deploying IDLV donor DNA and zinc-finger nuclease (ZFN) technology.
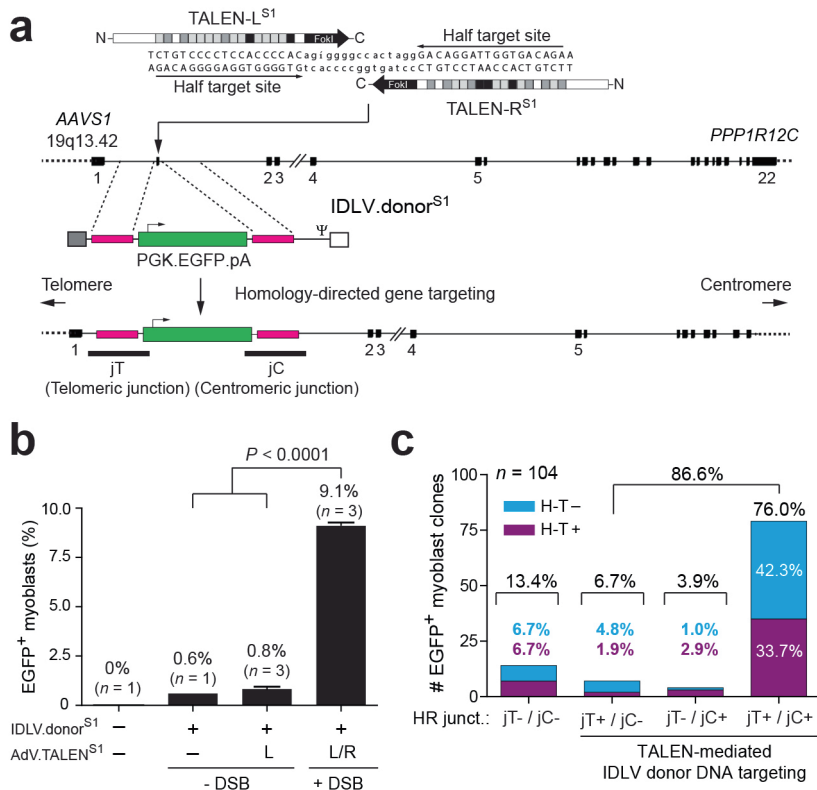

Figure 1. Gene targeting with IDLV donor DNA and TALEN-induced DSBs. (**a**) The schematic shows TALENs drawn in relation to their target sites in the *AAVS1* locus. These bipartite target sites frame spacer sequences, which are cleaved upon local nuclease pair dimerization. IDLV.donor[S1] vector genomes contain HR substrates consisting of a reporter expression unit flanked by sequences identical to those bracketing the TALEN target sequences. The transgene in IDLV.donor[S1] comprises the human *PGK1* promoter, the *EGFP* ORF and the bovine *GH1* polyadenylation signal. The structure of the integrated exogenous DNA resulting from error-free HR events, forming 'telomeric' and 'centromeric' junctions (jT and jC), is depicted. Ψ, HIV-1 packaging signal; white and gray boxes, 5′ and 3′ long terminal repeat, respectively. (**b**) Flow cytometric analysis on myoblasts cotransduced with the indicated viral vector constructs. Flow cytometry was performed at 27 d post transduction; 10,000 events, each corresponding to a single viable cell, were measured per sample. Error bars, s.d. *P* value (by two-tailed *t*-test) and the number of independent experiments (*n*) are shown. DSB, double-stranded DNA break. (**c**) Cumulative molecular characterization of EGFP+ myoblast clones generated by TALEN-L[S1]:TALEN-R[S1] and IDLV.donor[S1] DNA delivery (**Supplementary Fig. 1b**). The frequencies of EGFP+ myoblast clones with random insertions (jT−/jC−), HR-derived telomeric junctions (jT+/jC−), HR-derived centromeric junctions (jT−/jC+) and HR-derived telomeric and centromeric junctions (jT+/jC+) are plotted. The percentages of myoblast clones without (−) or with (+) head-to-tail IDLV concatemers (H-T) are also plotted.

Designer nuclease–induced AdV DNA targeting is accurate

Together with IDLVs, adeno-associated viral vectors (rAAVs) constitute the most commonly used viral vectors for the delivery of HR substrates into mammalian cells[4,19–22]. Like IDLVs, rAAV genomes have free ends and can become inserted at sporadic genomic DSBs after being co-opted by illegitimate recombination pathways involved in chromosomal DNA break repair[11,23]. In contrast, linear double-stranded AdV DNA has a terminal protein (TP) covalently attached at its 5′ ends. This led us to postulate that this capped DNA structure reduces the chance for interactions between the donor and off-target DSBs, such that pairing between acceptor and donor DNA through shared sequences would favor HR-dependent insertions at cleaved target sites. We therefore asked whether delivering donor DNA in the context of protein-capped AdV genomes[24] displays a less promiscuous chromosomal integration pattern and a more precise insertion profile than that resulting from using free-ended DNA. We generated the *E1*- and early region 2A (*E2A*)-deleted AdV.Δ2.donor[S1] to introduce the *AAVS1*-matched HR substrate into myoblasts exposed to the *AAVS1*-specific TALENs, as well as to unexposed cells. This AdV-delivered donor DNA harbors the same EGFP-encoding transcriptional unit present in the aforementioned IDLV.donor[S1] (**Fig. 1a**). Flow cytometry of the resulting long-term cultures showed a clear nuclease-dependent increase in the frequencies of stably transduced cells (**Fig. 2a**) with not only TALEN-induced but also residual DNA integration rates being lower than those measured in their IDLV.donor[S1]-transduced counterparts (**Fig. 1b**). The degree of the TALEN-dependent stimulatory effect was nonetheless similar to that observed in IDLV.donor[S1]-transduced myoblasts (**Fig. 1b**). Interestingly, cotransducing myoblasts with the AdV.Δ2.donor[S1] and TALEN-encoding AdVs resulted in EGFP+ populations with a distribution of transgene expression levels almost as narrow as those of clones harboring *AAVS1*-targeted donor[S1] DNA and much narrower than those of IDLV.donor[S1]-modified populations (**Fig. 2b**). We observed similar results at the clonal level by comparing the mean fluorescence intensity (MFI) of EGFP+ myoblasts randomly selected from TALEN-treated cultures transduced either with IDLV.donor[S1] or with AdV.Δ2.donor[S1] (**Fig. 2c**). Collectively, these data indicate scarce chromosomal positional effects on transgene activity in AdV.Δ2.donor[S1]-modified populations, possibly resulting from a preponderance of site-specific over random genomic DNA insertions. Indeed, all of the EGFP+ myoblast clones isolated from cultures cotransduced with AdV.Δ2.donor[S1] and TALEN-encoding AdVs (*n* = 110) had *AAVS1*–foreign DNA junctions resulting from HR events at both termini (**Fig. 2d** and **Supplementary Fig. 4a**).
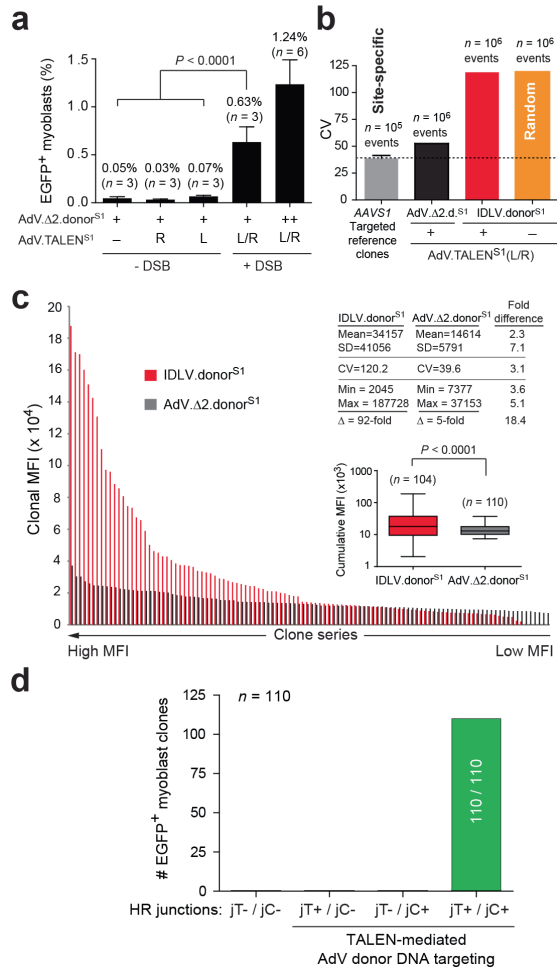
Figure 2. Gene targeting with AdV donor DNA and TALEN-induced DSBs. (**a**) Flow cytometric analysis on myoblasts cotransduced with the indicated viral vector constructs. The AdV.Δ2.donor[S1] was applied at multiplicity of infection values of 5 transducing units (TU)/cell (+) and 10 TU/cell (++). Cultures exposed to AdV.Δ2.donor[S1] at 5 TU/cell alone (+, −) or mixed with AdV.TALEN-L[S1] (+, L) or with AdV. TALEN-R[S1] (+, R) served as negative controls. Flow cytometry was performed at 45 d post transduction; 10,000 events corresponding to single viable cells were measured per sample. Error bars, s.e.m. P value (by two-tailed t-test) and the number of independent experiments (n) are shown. DSB, doubled-stranded DNA break. (**b**) Coefficient of variation (CV) of EGFP[+] populations exposed to *AAVS1*-specific TALENs and HR substrates delivered via the indicated donors. Clones containing an *AAVS1*-targeted donor[S1] DNA copy ($n = 10$ clones) and EGFP[+] populations with randomly inserted IDLV.donor[S1] DNA served as controls (gray and orange bars, respectively). (**c**) Mean fluorescence intensity (MFI) values of EGFP[+] myoblast clones targeted with TALENs and either IDLV.donor[S1] (red bars) or AdV.Δ2.donor[S1] (black bars). Inset, box plot of the cumulative MFI values corresponding to both series of myoblast clones analyzed; n indicates number of clones. Whiskers, minimum and maximum. P was calculated by two-tailed t-test. (**d**) Cumulative data on randomly selected EGFP[+] myoblast clones ($n = 110$) isolated from cultures cotransduced with AdV.Δ2.donor[S1], AdV.TALEN-L[S1] and AdV.TALEN-R[S1] (**Supplementary Fig. 4a**). The frequencies of EGFP[+] myoblast clones representing the different integrant classes are plotted as in **Figure 1c**.

To further probe the precision of AdV gene targeting, we set up a PCR assay to detect head-to-tail exogenous DNA concatemers. This assay failed to produce any discernible head-to-tail–specific PCR species from genomic DNA of EGFP+ myoblasts sorted from cultures cotransduced with AdV.Δ2.donor[S1] and TALEN-encoding AdVs (**Supplementary Fig. 4b**). Southern blot analysis of AdV.Δ2.donor[S1]-modified clones confirmed target-site specificity. We note that we did identify a clone that, in addition to the typical *AAVS1*-targeted donor DNA, contained an integrant whose origin is consistent with HR-independent integration (**Supplementary Fig. 5a**). Nonetheless, in line with its low prevalence, we did not detect the latter type of integrant in the parental EGFP+ myoblast population (**Supplementary Fig. 5b**). Southern blot analysis also identified a clone that underwent biallelic targeting.

Next we performed AdV gene-targeting experiments in HeLa cells. These cells display a high degree of genetic instability (**Supplementary Fig. 6**), providing, as a result, a more stringent model system in which to evaluate HR-mediated genome editing amidst a presumably high frequency of spontaneous chromosomal DSBs. Consistent with the experiments carried out in myoblasts (**Fig. 2a**), we observed a significant ($P < 0.0001$) TALEN-dependent increase in the frequency of stably transduced cells (**Fig. 3a,b**). We did not detect head-to-tail AdV DNA concatemers in the EGFP+ populations resulting from TALEN-induced chromosomal insertion of Adv.Δ1.donor[S1] DNA (**Fig. 3c**). Notably, all randomly selected EGFP+ HeLa cell clones ($n = 83$) were genetically modified through homology-directed gene targeting at *AAVS1*, confirming a high level of target-site specificity following AdV-mediated delivery of donor DNA (**Fig. 3d** and **Supplementary Fig. 7**).
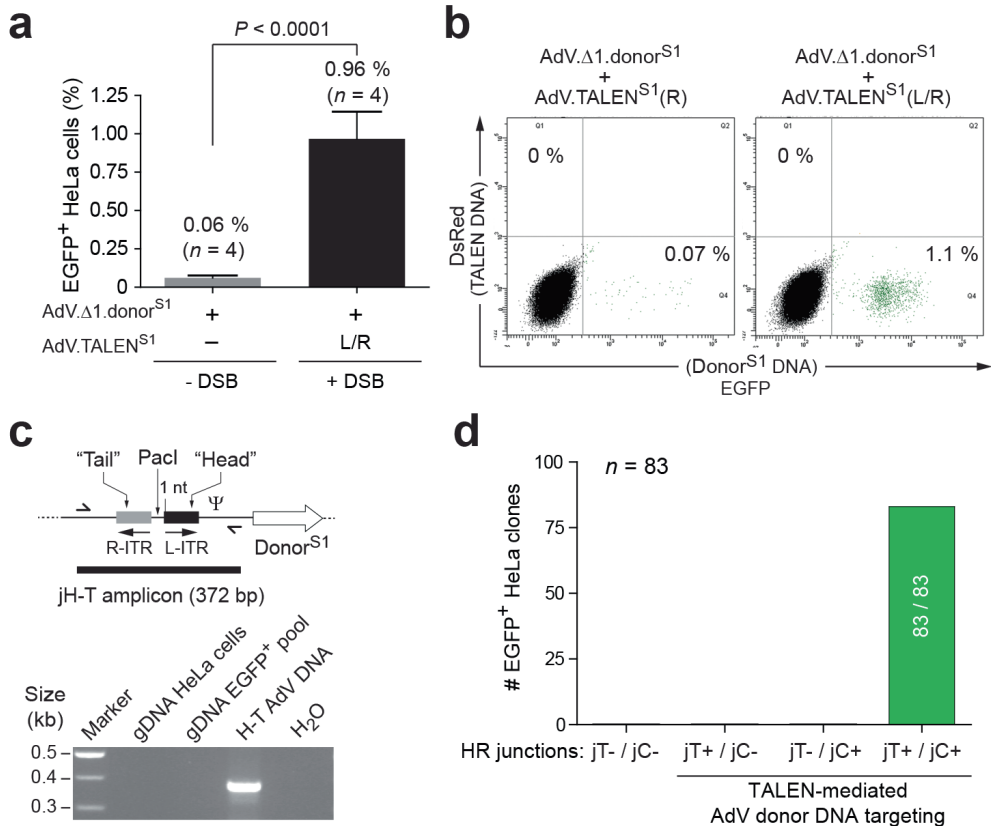
Figure 3. Nuclease-mediated gene targeting of AdV donor DNA in genetically unstable HeLa cells. (**a**) Flow cytometric analysis of HeLa cells 24 d after transduction with the indicated viral vector constructs. Error bars, s.d. P value (by two-tailed t-test) and the number of independent experiments (n) are shown. (**b**) Representative flow cytometry dot plots corresponding to HeLa cell cultures exposed to the indicated viral vector constructs at 24 d post transduction. Owing to a bicistronic expression unit, cells transduced with TALEN-encoding AdVs become tagged with DsRedEx2.1. (**c**) PCR analysis probing for head-to-tail AdV DNA concatemers in transduced EGFP+ HeLa cells. Top, schematics of *in vitro*–assembled head-to-tail AdV DNA junctions (jH-T). L-ITR and R-ITR, 'left' and 'right' AdV ITR, respectively; half arrows, primers; horizontal bar, amplicon diagnostic for the presence of head-to-tail AdV DNA concatemers. Bottom, PCR analysis of genomic DNA from EGFP+ HeLa cells exposed to *AAVS1*-specific TALENs and AdV.Δ1.donor^{S1} (gDNA EGFP+). DNA from parental HeLa cells (gDNA HeLa cells) and nuclease-free water are negative controls; *in vitro*–generated head-to-tail AdV DNA (H-T AdV DNA) is the positive control. (**d**) Cumulative data on randomly selected EGFP+ HeLa cells isolated from cultures co-transduced with AdV.Δ1.donor^{S1}, AdV.TALEN-L^{S1} and AdV.TALEN-R^{S1} (**Supplementary Fig. 7**). The frequencies of EGFP+ clones representing the different integrant classes are plotted as in **Figure 1c**.

AdV-based gene targeting with CRISPR-Cas9

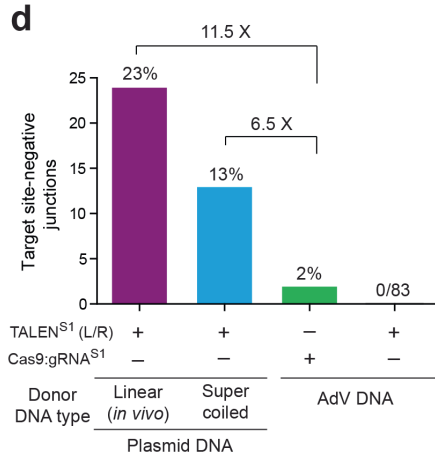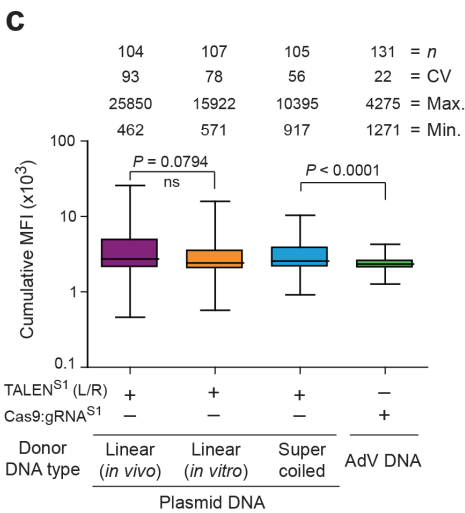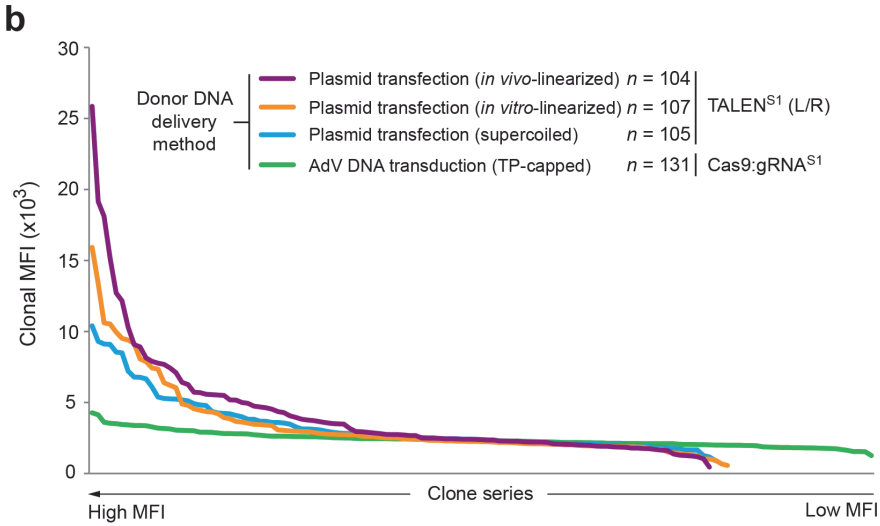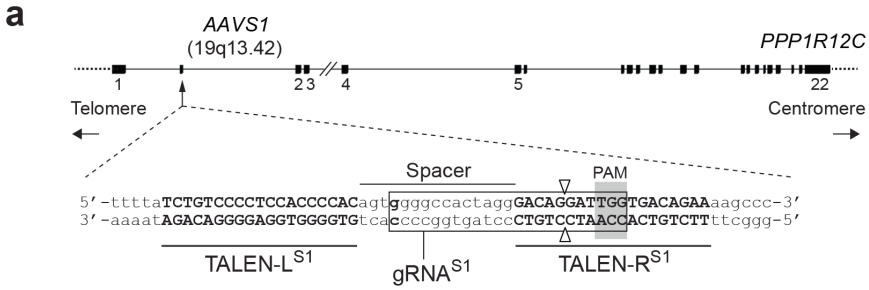We investigated the compatibility of nuclease-induced AdV gene targeting with the versatile CRISPR-Cas9 RNA-guided nuclease system[25]. We deployed the vector pair AdV.Cas9 and AdV.gRNA[S1], which encode Cas9 and a single guide RNA (gRNA[S1]) addressing the Cas9 nuclease to a genomic position overlapping with the target site for the *AAVS1*-specific TALENs (**Fig. 4a**). Exposing HeLa cells to AdV.Cas9 and AdV.gRNA[S1] resulted in robust and dose-dependent DSB formation at *AAVS1* (**Supplementary Fig. 8**).

We initiated gene-targeting experiments by cotransducing HeLa cells with AdV.Δ2.donor[S1], AdV.Cas9 and AdV.gRNA[S1]. HeLa cells cotransduced exclusively with AdV.Δ2.donor[S1] and AdV.Cas9 served as negative controls. In these experiments, we also tested conventional nonviral plasmid donors with both circular and linear topologies deploying, in this case, TALENs as the designer nuclease system. Thus, HeLa cells were cotransfected with constructs encoding the *AAVS1*-specific TALENs mixed with supercoiled pAdV.donor[S1], *Pac*I-linearized pAdV.donor[S1] or TALEN-cleavable pAdV.donor[S1/T-TS] (**Supplementary Fig. 9**). The free-ended plasmid donor templates were generated *in vitro* and *in cellula* by restriction enzyme– and TALEN–induced DNA cleavage, respectively. Of note, the nonprokaryotic DNA portions of these plasmids are isogenic to those present in protein-capped AdV.Δ2.donor[S1] genomes. HeLa cells transfected with the TALEN-L[S1]-expressing construct plus each donor plasmid type served as negative controls.

The percentages of EGFP[+] cells in cultures exposed to Cas9:gRNA[S1] complexes and AdV.Δ2.donor[S1] (**Supplementary Fig. 9**) were comparable to those measured in populations treated with TALEN-L[S1]:TALEN-R[S1] dimers and AdV.Δ2.donor[S1] (**Fig. 3a,b**). Moreover, the frequencies of genetically modified cells were clearly higher in cultures that had been subjected to site-specific DSBs. Crucially, stably transduced HeLa cells generated by co-delivering CRISPR-Cas9 complexes and AdV.Δ2.donor[S1] DNA displayed a remarkably narrow range of transgene activities, as determined by flow cytometric screening of randomly selected EGFP[+] clones (**Fig. 4b,c**), reminiscent of data in myoblasts modified with TALENs and AdV donor DNA (**Fig. 2c**).

Figure 4. Nuclease-induced gene targeting with AdV- versus plasmid-mediated delivery of HR substrates. (**a**) *AAVS1* recognition sequences for the TALEN and Cas9:gRNA complexes are drawn in relation to the *PPP1R12C* locus in which they are embedded. The target sites of the TALEN pair and the RNA-guided CRISPR-Cas9 nuclease are shown in upper case and boxed, respectively. The protospacer-adjacent motif (PAM) is shaded. Open vertical arrowheads indicate the Cas9 cleavage site. (**b**) MFI distribution of randomly selected EGFP[+] HeLa cell clones genetically modified by Cas9:gRNA[S1] and AdV.Δ2.donor[S1] (TP capped) or by TALEN-L[S1]:TALEN-R[S1] (L/R) and either pAdV.donor[S1] (supercoiled), *Pac*I linearized pAdV.donor[S1] (*in vitro* linearized) or pAdV.donor[S1/T-TS] (*in vivo* linearized). (**c**) Box plot of cumulative MFI values corresponding to panel **b**. Whiskers, minimum and maximum; ns, not significant (two-tailed *t*-test), *n* indicates number of clones. (**d**) Frequencies of EGFP[+] clones lacking the *AAVS1*-donor DNA 'telomeric' junction, as identified by PCR screening of HeLa cell clones (**Supplementary Figs. 7** and **11**). The data correspond to the clones analyzed in **b,c** and **Supplementary Figure 7**. ▶

a

AAVS1
(19q13.42)

PPP1R12C

Telomere

Centromere

Spacer   PAM

5'-ttttaTCTGTCCCCTCCACCCCACagtggggccactaggGACAGGATTGGTGACAGAAaagccc-3'
3'-aaaatAGACAGGGGAGGTGGGGTGtcacccggtgatccCTGTCCTAACCACTGTCTTttcggg-5'

TALEN-L^{S1} gRNA^{S1} TALEN-R^{S1}

b

Donor DNA delivery method

— Plasmid transfection (*in vivo*-linearized) *n* = 104
— Plasmid transfection (*in vitro*-linearized) *n* = 107   TALEN^{S1} (L/R)
— Plasmid transfection (supercoiled)   *n* = 105
— AdV DNA transduction (TP-capped)   *n* = 131   Cas9:gRNA^{S1}

Clonal MFI ($\times10^3$)

Clone series

High MFI   Low MFI

6

c

104	107	105	131 = *n*
93	78	56	22 = CV
25850	15922	10395	4275 = Max.
462	571	917	1271 = Min.

Cumulative MFI ($\times10^3$)

P = 0.0794   ns

P < 0.0001

TALEN^{S1} (L/R)	+	+	+	−
Cas9:gRNA^{S1}	−	−	−	+
Donor DNA type	Linear (*in vivo*)	Linear (*in vitro*)	Super coiled	AdV DNA
		Plasmid DNA		

d

Target site-negative junctions

11.5 X

6.5 X

23%

13%

2%   0/83

TALEN^{S1} (L/R)	+	+	−	+
Cas9:gRNA^{S1}	−	−	+	−
Donor DNA type	Linear (*in vivo*)	Super coiled	AdV DNA	
	Plasmid DNA			

In contrast, cells exposed to TALENs and plasmid donors led to EGFP[+] clones displaying significantly ($P < 0.0001$) broader distributions of transgene expression levels, independently of the topology of the donor DNA (**Fig. 4b,c**). The higher homogeneity of transgene expression among AdV-modified cells is also apparent in the correlation between the coefficient of variation (CV) and MFI values for each individual clone (**Supplementary Fig. 10a**), as well as by comparing the range of transgene expression levels in EGFP[+] populations resulting from CRISPR-Cas9–induced AdV gene targeting to those obtained using plasmid donors as templates (**Supplementary Fig. 10b**).

PCR analysis of donor DNA-*AAVS1* junctions demonstrated higher target-site specificity of AdV over plasmid HR substrates even when using the presumably more promiscuous[26–28], yet more versatile, CRISPR-Cas9 system[25] (**Fig. 4d** and **Supplementary Fig. 11**). Furthermore, as previously observed for IDLV donor DNA, plasmid donor DNA resulted in intermolecular illegitimate recombination events (**Supplementary Fig. 11a**). Of note, these data represent, to the best of our knowledge, the first demonstration of the utility of the CRISPR-Cas9 system for viral vector–mediated exogenous DNA targeting in human cells.

There are precedents for the delivery and persistence of bacterial DNA in mammalian cells exposed to both nonviral and viral vectors[29,30]. Besides their unpredictable structures, these prokaryotic DNA footprints are also undesirable owing to their immunostimulatory and methylation-prone nucleotide patterns (for example, CpG motifs). We probed for bacterial DNA in cell populations genetically modified by AdV- and plasmid-based gene targeting, and we detected *Kan*[R] DNA exclusively in the latter case (**Supplementary Fig. 12**).

Capped AdV donor DNA reduces illegitimate recombination

Finally, we tested whether the protein cap of AdV genomes is a determinant of the specificity and fidelity of the AdV gene-targeting process. To this end, we generated AdV.Δ2.donor[S1/T-TS] and used it together with AdV.Δ2.TALEN-L[S1] and AdV.Δ2.TALEN-R[S1] to transduce HeLa cells. The AdV.Δ2.donor[S1/T-TS] has the same genetic makeup as the AdV.Δ2.donor[S1] except that its donor DNA payload is flanked by recognition sequences for the *AAVS1*-specific TALENs (T-TS) (**Fig. 5a**). The AdV.Δ2.TALEN-L[S1] and AdV.Δ2.TALEN-R[S1] differ from AdV.TALEN-L[S1] and AdV.TALEN-R[S1], respectively, in that they are deleted simultaneously in *E1* and *E2A*. In striking difference with the results obtained with the T-TS–negative AdV (**Fig. 3d** and **Supplementary Fig. 7**), PCR screening of thirty EGFP[+] HeLa cell clones expanded from cultures exposed to TALENs and AdV.Δ2.donor[S1/T-TS] revealed that seven of these lines lacked HR-derived *AAVS1*–exogenous DNA centromeric junctions (**Fig. 5a**). Control experiments established the release of donor[S1/T-TS] DNA from the AdV.Δ2.donor[S1/T-TS] genome in transduced cells (**Supplementary Fig. 13**). These results confirm that HR substrates delivered in the context of protein-capped AdV genomes lead to more precise gene targeting when compared to HR substrates transferred in free-ended linear templates (**Fig. 5b**). We obtained similar results following

AdV transduction experiments in reporter cells (**Supplementary Results** and **Supplementary Fig. 14**) and differentiating myoblasts (**Supplementary Results** and **Supplementary Fig. 15**).
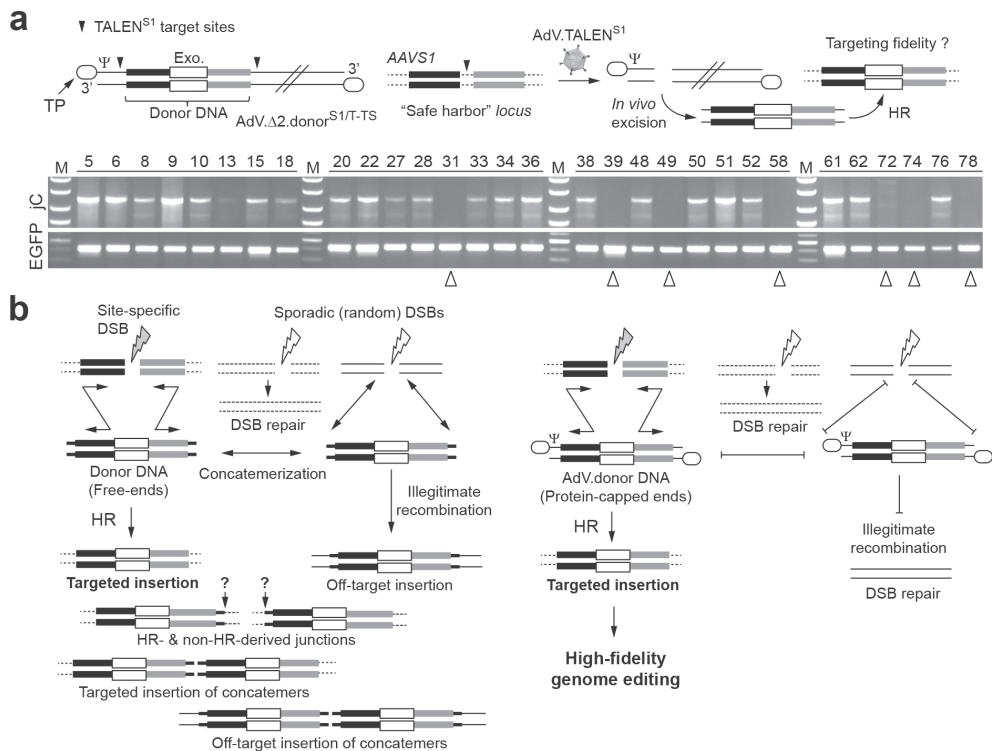

Figure 5. The protein cap of AdVs contributes to gene-targeting specificity and accuracy. (**a**) The effect of donor DNA excision from the protein-capped AdV genomes on gene targeting frequency. Top, experimental strategy for the TALEN-mediated excision of HR templates from AdV DNA in transduced cells. AdV.Δ2.donor$^{S1/T-TS}$, AdV carrying donor^{S1} DNA framed by target sequences for the *AAVS1*-specific TALENs (solid vertical arrowheads); Open oval, terminal protein (TP) covalently attached to the 5′ termini of AdV DNA; Ψ, AdV packaging signal; Exo., exogenous DNA. Bottom, PCR screening of EGFP$^+$ HeLa cell clones to detect DNA junctions formed by HR events between AdV.Δ2.donor$^{S1/T-TS}$ DNA and the 'centromeric' side of the *AAVS1* locus. Open arrowheads, clones harboring integrants lacking HR-derived centromeric junctions. (**b**) Model for high-fidelity genome editing based on site-specific DSBs and AdV-delivered donor DNA. Left, donor DNA with free ends are sensed as broken DNA and rerouted to off-target or sporadic DSBs. Precision can be further compounded by the chromosomal insertion of concatemeric vector DNA forms. Right, protein-capped AdV genomes make interactions between donor templates and off-target or sporadic DSBs less probable.

DISCUSSION

The overall performance of genome-editing technology depends on its efficiency as well as on its specificity and accuracy. On-target specificity is being actively investigated, with significant efforts devoted to not only examining but also reducing off-target DNA cleaving activities of artificial nucleases[5,26–28,31–34]. The contribution of the donor DNA component to the specificity and accuracy with which different exogenous HR substrates are inserted at a target site has, comparatively, received scarcer scrutiny. It is known, however, that commonly used viral vector–borne donor templates lead to off-targeted events, presumably owing to their capture by random chromosomal DSBs[11,23,29]. In fact, IDLV and rAAV genomes are being exploited as tags for mapping the distribution of DSBs at a genome-wide scale[11,35,36]. Experimental evidence indicates that linear DNA molecules, by resembling DSB repair intermediates, are targets for illegitimate recombination mechanisms regardless of their viral or nonviral origin[37,38]. The ensuing processing of these molecules by DNA repair factors can result in their off-target chromosomal insertion as single or multiple copies. These outcomes reduce the predictability of transgene expression and gene repair in target cell populations. In principle, viral vector–mediated genome editing might be refined by adding, 'outside' the targeting HR module, a heterologous cassette conferring negative selection against cells with insertions at off-target sites. However, purging cells with targeted insertions formed by a combination of nonhomologous end joining and HR and/or with irregular concatemeric footprints would be less certain. Besides being complex and measurably 'leaky', negative selection strategies also reduce the effective packaging capacity of vector particles. The overall fidelity of the IDLV-based genome-editing process may also suffer from the intrinsic genetic 'plasticity' of lentiviruses. For instance, the error rates of retroviral reverse transcriptases are orders of magnitude higher than those of dsDNA viral polymerases[39].

The frequencies of genetic modification achieved by DSB-induced AdV gene targeting were lower than those obtained with the IDLV platform but were within the range of or above those reached with plasmid donors co-delivered with nucleases or reported for single-stranded rAAV templates[40]. Regardless, the high packaging capacity of AdV particles coupled to the availability of robust positive selection strategies can, if necessary, be exploited to rapidly enrich for the AdV-modified population. Notably, the precision of the AdV gene-targeting procedure is expected to reduce the dependency on time-consuming screening of genetically modified populations to identify properly targeted cells.

On the basis of our findings, we put forward the view that the numerous efforts devoted to minimizing off-target activity of sequence-specific nucleases[5,31–34] should be complemented with those aiming at identifying HR substrates whose features maximize on-target and accurate insertion of foreign DNA. The development of these optimized HR templates is expected to promote fundamental and applied research activities dependent on the precise manipulation of mammalian genomes.

METHODS

Cells. The HeLa cells (American Type Culture Collection) and EGFP-positive H27 clone derivative[41] were cultured in Dulbecco's modified Eagle's medium (DMEM; Invitrogen) supplemented with 5% fetal bovine serum (FBS; Invitrogen). The HEK 293T lentiviral vector producer cells were maintained in DMEM containing 10% FBS, whereas the AdV packaging cell lines PER.C6 (ref. 42) and PER.E2A[43] were cultured in DMEM supplemented with 10% FBS and 10 mM $MgCl_2$ (Sigma-Aldrich) in the absence and in the presence of 250 µg/ml of Geneticin (Invitrogen), respectively. These cell types were kept in a humidified atmosphere containing 10% CO2. The origin of and the culture conditions for the human myoblasts have been detailed elsewhere[44,45]. The cell batches used to generate the viral vector preparations and to carry out the experiments were mycoplasma free.

Recombinant DNA. The complete and annotated DNA sequences of lentiviral vector shuttle plasmids AP45_pLV.donor[EGFP] and AQ25_pLV.donor[S1] can be retrieved via GenBank accession numbers KF419293 and KF419294, respectively. The AdV shuttle plasmid AQ60_pAdV.donor[EGFP] contains as foreign DNA the same targeting module as that in AP45_pLV.donor[EGFP]. The AdV shuttle plasmid pAdV.donor[S1/T-TS] was constructed by inserting 'upstream' and 'downstream' of the donor[S1] DNA module in pSh.AAVS1.eGFP[14] (herein named pAdV.donor[S1]) two annealed oligodeoxyribonucleotides containing bipartite target sequences for the *AAVS1*-specific TALENs (T-TS). A similar approach based on inserting into pAdV.donor[S1] a direct repeat of *FRT* sites in place of the T-TS sequences was pursued in parallel. These maneuvers resulted in pAdV.donor[S1/FRT]. The AdV molecular clones AL25_pAdV.ΔE1.donor[S1].F[50] and AL27_pAdV.ΔE1ΔE2A.donor[S1].F[50] were assembled by HR in *Escherichia coli* strains[46] BJ5183[pAdEasy-1.50] and BJ5183[pAdEasy-2.50], respectively, transformed with *Mss*I-treated pAdV.donor[S1]. The AdV molecular clones pAdV.ΔE1ΔE2A.donor[S1/T-TS].F[50] and pAdV.ΔE1ΔE2A.donor[S1/FRT].F[50] were built by HR following the transformation of the latter cells with *Mss*I-digested plasmids pAdV.donor[S1/T-TS] and pAdV.donor[S1/FRT], respectively. The FP635-encoding AdV molecular clone AR02_pAdV.ΔE1.ΔE2A.donor[EGFP].F[50] was generated by HR in BJ5183[pAdEasy-2.50] cells[46] transformed with *Mss*I-digested AQ60_pAdV.donor[EGFP]. The AdV shuttle plasmid AG03_pAdV.CMV.TALEN-L[EGFP] and AF59_pAdV.CMV.TALEN-R[EGFP] were constructed by inserting into the multiple cloning site of AQ17_pAdV.MCS.SV40pA the CMV promoter linked to the ORFs of TALEN-L[EGFP] and TALEN-R[EGFP], respectively. The TALEN-L[EGFP] and TALEN-R[EGFP] proteins (**Supplementary Fig. 14a**) were custom designed (GeneArt) to recognize the EGFP half target sites 5′-TGAACTTCAAGATCCGCCA-3′ and 5′-TCGGCGAGCTGCACGCTGC-3′, respectively, and cleave within their 14-bp intervening spacer sequence (**Supplementary Fig. 14a**). Next, the full-length AdV molecular clones AF50_pAdV.ΔE1.TALEN-L[EGFP].F[50] and AF52_pAdV.ΔE1.TALEN-R[EGFP].F[50] were assembled by HR in *E. coli* by transforming strain BJ5183[pAdEasy-1.50] with *Mss*I-treated AdV shuttle plasmids AG03_pAdV.CMV.TALEN-L[EGFP] and AF59_pAdV.CMV.TALEN-R[EGFP], respectively.

The AdV shuttle plasmid pAdSh.PGK.Cas9 contains a humanized ORF encoding the *Streptococcus pyogenes* nuclease Cas9 under the transcriptional control of the *PGK1* promoter and the SV40 polyadenylation signal, whereas the AdV shuttle plasmid pAdSh.U6.gRNA^{S1} encodes a *U6* promoter–driven single guide RNA targeting the Cas9 protein to the human *AAVS1* locus[47]. The human codon-optimized *cas9* ORF and the RNA Pol III-dependent *gRNA^{S1}* expression unit have been published elsewhere[48] and were isolated from constructs hCas9 (Addgene plasmid 41815) and gRNA_AAVS1-T2 (Addgene plasmid 41818), respectively. Next, the AdV molecular clones pAdV^{Δ2}P.Cas9.F^{50} and pAdV^{Δ2}U6.gRNA^{S1}.F^{50} were generated by HR in BJ5183$^{pAdEasy-2.50}$ cells[46] following their transformation with *Mss*I-treated pAdSh.PGK.Cas9 and pAdSh.U6.gRNA^{S1}, respectively[47]. Further details about DNA constructs generated for this study can be found in the **Supplementary Note**.

Production and titration of AdV vectors. The generation of the fiber-modified *E1*-deleted AdVs Ad.ΔE1.TALEN-L^{S1}.F^{50} and Ad.ΔE1.TALEN-R^{S1}.F^{50} (herein referred to as AdV.TALEN-L^{S1} and AdV.TALEN-R^{S1}, respectively) has been detailed elsewhere[46]. The same applies to the fiber-modified *E1*- and *E2A*-deleted AdVs Ad.ΔE1ΔE2A.TALEN-L^{S1}.F^{50} and Ad.ΔE1ΔE2A.TALEN-R^{S1}.F^{50} (herein named AdV.Δ2.TALEN-L^{S1} and AdV.Δ2.TALEN-R^{S1}, respectively)[16]. The productions of the fiber-modified *E1*-deleted AdV AdV.Δ1.donor^{S1} and of its *E1*- plus *E2A*-deleted derivative AdV.Δ2.donor^{S1}, were initiated by transfecting PER.C6 and PER.E2A cells with *Pac*I-linearized AL25_pAdV.ΔE1.donor^{S1}.F^{50} and AL27_pAdV.ΔE1ΔE2A.donor^{S1}.F^{50}, respectively. The productions of the fiber-modified, EGFP-encoding, donor AdVs AdV.Δ2.donor$^{S1/T-TS}$ and AdV.Δ2.donor$^{S1/FRT}$ as well as that of the fiber-modified, *E1*- and *E2A*-deleted, FP635-encoding, donor AdV AdV.Δ2.donor^{EGFP} were started by transfecting PER.E2A cells with *Pac*I-treated pAdV.ΔE1ΔE2A.donor$^{S1/T-TS}$.F^{50}, pAdV.ΔE1ΔE2A.donor$^{S1/FRT}$.F^{50} and AR02_pAdV.ΔE1.ΔE2A.donor^{EGFP}.F^{50}, respectively. The *E1*- plus *E2A*-complementing PER.E2A cells were also used for rescuing and propagating the fiber-modified AdVs AdV.Cas9 and AdV.gRNA^{S1} following their transfection with the *Pac*I-linearized molecular clones pAdVΔ2P.Cas9.F^{50} and pAdVΔ2U6.gRNA^{S1}.F^{50}, respectively. Finally, the generation of the first-generation and fiber-modified AdVs AdV.TALEN-L^{EGFP} and AdV.TALEN-R^{EGFP}, was carried out in *E1*-only complementing cells transfected with *Pac*I-digested AF50_pAdV.ΔE1.TALEN-L^{EGFP}.F^{50} and AF52_pAdV.ΔE1.TALEN-R^{EGFP}.F^{50}, respectively.

The DNA transfection–mediated rescue of AdV particles in packaging cell lines and their subsequent propagation and purification were performed essentially as described previously[46,49]. The isolation and restriction fragment length analysis of AdV.Δ2.donor$^{S1/T-TS}$ and AdV.Δ2.donor$^{S1/FRT}$ DNA was carried out as detailed elsewhere[46,49]. The titers of the various reporter-encoding AdV stocks, expressed in terms of transducing units (TU) per ml, were determined through limiting dilutions on HeLa indicator cells seeded at a density of 8×10^4 cells per well of 24-well plates. At 3 d post transduction, frequencies of reporter-positive cells were measured by reporter-directed flow cytometry. The titers of the reporter-

negative AdV preparations were established by TCID$_{50}$ assays in complementing cells and by fluorometric quantification of genome-containing vector particles (VP) per milliliter as described elsewhere[46,49].

Production and titration of lentiviral vectors. The generation of the vesicular stomatitis virus glycoprotein G (VSV-G)-pseudotyped lentiviral vectors LV.ZFN-1^{EGFP} and LV.ZFN-2^{EGFP} has been described before[50]. The generation of the VSV-G-pseudotyped vectors IDLV.donor^{EGFP} and IDLV.donor^{S1} was carried out by transient transfections of 293T cells with shuttle plasmids AP45_pLV.donor^{EGFP} and AQ25_pLV.donor^{S1}, respectively, together with packaging construct51 AM16_psPAX2.IN^{D116N} and pseudotyping construct pLP/VSVG (Invitrogen) as detailed elsewhere[51,52]. The physical particle titers of lentiviral vector preparations were determined by using the RETRO-TEK HIV-1 p24 ELISA kit as specified by the manufacturer (Gentaur Molecular Products). Titers of these vector stocks expressed in terms of TU/ml were derived by using a conversion factor of 2,500 TU/ng of HIV-1 p24^{gag} protein.

Transduction experiments. The IDLV.donor^{EGFP} transductions on cultures of EGFP$^+$ H27 indicator cells were carried out as follows. Eighty-thousand cells were seeded in wells of 24-well plates (Greiner Bio-One). The next day, IDLV.donor^{EGFP} was added at a multiplicity of infection (MOI) of 45 TU/cell together with the LV.ZFN-1^{EGFP} and LV.ZFN-2^{EGFP} vectors each applied at an MOI of 8 TU/cell. Parallel H27 cultures that were either untreated or were incubated exclusively with IDLV.donor^{EGFP} at an MOI of 45 TU/cell, served as controls. Next, after an extensive 5-week subculturing period, the frequencies of reporter-positive and reporter-negative H27 cell populations were monitored and quantified, respectively, by two-color fluorescence microscopy and flow cytometry.

Long-term transduction experiments on human myoblasts were initiated by seeding 2 × 10^5 cells per well in 24-well plates. The next day, the cells were incubated with AdV.TALEN-L^{S1} (2.5 TU/cell) and AdV.TALEN-R^{S1} (2.5 TU/cell) mixed with IDLV.donor^{S1} (10 TU/cell) or with AdV.Δ2.donor^{S1} (10 TU/cell). The frequencies of IDLV.donor^{S1}- and AdV.Δ2.donor^{S1}-modified myoblasts were determined at 27 and 45 d post transduction, respectively.

Long-term transduction experiments on HeLa cells were started by seeding 8 × 10^4 cells per well in 24-well plates. After an overnight incubation period, the cells were exposed to AdV.TALEN-L^{S1} (3 TU/cell) and AdV.TALEN-R^{S1} (3 TU/cell) together with AdV.Δ1.donor^{S1} (6 TU/cell). The frequencies of stably transduced cells were measured by flow cytometry at 24 d post transduction.

To investigate the effect of TALEN-mediated donor DNA excision on the rate of gene targeting, 8 × 10^4 HeLa cells, seeded 1 d before, were transduced with AdV.Δ2.TALEN-L^{S1} (1.5 TU/cell) and AdV.Δ2.TALEN-R^{S1} (1.5 TU/cell) together with AdV.Δ2.donor$^{S1/T-TS}$ (3 TU/cell). The frequencies of stably transduced cells were established at 24 d post transduction by flow cytometry. The EGFP$^+$ populations present in the various long-term cultures were sorted, after which over 250 individual clones were randomly selected for expansion and molecular analyses.

The gene-targeting experiments on EGFP[+] H27 cells with the FP635-encoding vector AdV.Δ2.donor[EGFP] were carried out as follows. Two-hundred thousand cells were seeded in wells of 24-well plates. The next day, AdV.Δ2.donor[EGFP] was added at an MOI of 4 TU/cell together with 1:1 mixtures of AdV.TALEN-L[EGFP] and AdV.TALEN-R[EGFP] applied at total doses of 6.0×10^3, 2.4×10^3 and 9.6×10^2 VP/cell. Controls consisted of H27 cells, HeLa cells exposed for 3 days to AdV.Δ2.donor[EGFP] (0.5 TU/cell) and H27 cells transduced with AdV.Δ2.donor[EGFP] (4 TU/cell) plus AdV.TALEN-L[EGFP] at 6.0×10^3, 2.4×10^3 and 9.6×10^2 VP/cell. At 48 h post transduction, the various H27 cultures received fresh regular medium, and 1 d later they started to be subcultured at a rate of about twice per week in order to remove the episomal vector DNA. The frequencies of the various types of H27 cell populations were determined by two-color flow cytometry at 37 d post transduction.

Flow cytometry and light microcopy. The measurement of transgene expression parameters (i.e., frequencies of reporter-positive and reporter-negative target cells, mean fluorescence intensities and coefficients of variation) were determined by using a BD LSR II flow cytometer (BD Biosciences). Data were analyzed with the aid of BD FACSDiva 6.1.3 software (BD Biosciences) or FlowJo 7.2.2 (Tree Star). Mock-transduced target cells were used to set background fluorescence levels. At least 10,000 viable single cells were analyzed per sample. The light microscopic analyses were carried out with an IX51 inverse fluorescence microscope equipped with a XC30 Peltier-cooled digital color camera (Olympus). The images were processed with the aid of CellF 3.4 imaging software (Olympus).

Functional validation of CRISPR-Cas9 complexes delivered into target cells by AdVs. Fifty-thousand HeLa cells were transduced with AdV.Cas9 alone (300 TCID$_{50}$/cell) or with 1:1 mixtures of AdV.Cas9 and AdV.gRNA[S1] applied at total vector doses of 60, 120, 180, 240 and 300 TCID$_{50}$/cell. At 3 d post transduction, chromosomal DNA was extracted from the target cells by using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. Next, *AAVS1*-specific PCR amplifications and T7 endonuclease I–based genotyping assays were carried out essentially as previously described[49].

Designer nuclease–mediated gene targeting of linear and supercoiled plasmid donor DNA. HeLa cells were seeded at a density of 6.5×10^4 cells per well of 24-well plates. The next day, the cells were transfected with DNA mixtures consisting of 100 ng of 1383.pVAX.AAVS1.TALEN.L-94 (ref. 16), 100 ng of 1384.pVAX.AAVS1.TALEN.R-95 (ref. 16) and 200 ng of *AAVS1*-targeting AdV donor DNA plasmids. The targeting constructs were pAdV.donor[S1], *Pac*I-linearized pAdV.donor[S1] and pAdV.donor[S1/T-TS]. Controls were provided by transfecting HeLa cells with 200 ng of 1383.pVAX.AAVS1.TALEN.L-94 mixed together with 200 ng of pAdV.donor[S1], *Pac*I-linearized pAdV.donor[S1] or pAdV.donor[S1/T-TS]. The completion of the *Pac*I digestions was confirmed by agarose gel electrophoreses and ethidium bromide staining. Each of the plasmid mixtures were diluted in 50 µl of 150 mM NaCl and received 1.32 µl of a 1 mg/ml polyethylenimine solution under vigorous shaking for about 10 s. After a 20-

min incubation period at room temperature, the resulting polycation-DNA complexes were directly added into the culture medium. After 7 h, the transfection mixtures were removed and fresh culture medium was added. The resulting cell populations were subsequently subjected to subculturing for 3 weeks, after which cells stably expressing EGFP in these populations were individually sorted by flow cytometry into wells of 96-well plates. Viable single cell-derived clones corresponding to the various experimental settings were randomly selected for transgene expression and integration status analysis.

Gene targeting of AdV donor DNA by using the RNA-guided nuclease Cas9. HeLa cells were seeded at a density of 8.0×10^4 cells per well in 24-well plates. The following day, the cells were transduced with AdV.Cas9 (150 TCID$_{50}$/cell), AdV.gRNA[S1] (50 TCID$_{50}$/cell) and AdV.Δ2.donor[S1] (10 TU/cell). To serve as negative controls, HeLa cells were either mock-transduced or were transduced with AdV.Cas9 (150 TCID$_{50}$/cell) and AdV.Δ2.donor[S1] (10 TU/cell). At 3 d post transduction, mock- and vector-transduced HeLa cells started to be subcultured. At 17 d post transduction, EGFP stably expressing cells were individually sorted by flow cytometry into wells of 96-well plates. Viable single cell-derived clones isolated from cultures initially exposed to AdV.Cas9, AdV.gRNA[S1] and AdV.Δ2.donor[S1] were randomly selected for transgene expression and integration status analysis.

Cell sorting and clonal expansion. Flow cytometry–assisted cell sorting was done after the removal of donor DNA-containing episomes from long-term HeLa and human myoblast cultures. The EGFP$^+$ cells were collected in 1:1 mixtures of regular medium containing 2× penicillin-streptomycin (Invitrogen) and FBS. Next, the sorted cells were individually seeded in wells of 96-well plates (Greiner Bio-One) at a density of 0.3 cells per well in their respective medium supplemented with 50 μM of α-thioglycerol (Sigma-Aldrich) and 20 nM of bathocuprione disulphonate (Sigma-Aldrich) to increase cloning efficiency[53]. Finally, over 250 individual clones were randomly selected for expansion and molecular analyses.

Genomic DNA extraction. Genomic DNA was extracted from cell populations and clones essentially as described before[54]. In brief, the cells were collected and incubated overnight at 55 °C in 500 μl of lysis buffer (100 mM Tris-HCl, pH 8.5, 5 mM EDTA, 0.2% sodium dodecyl sulfate and 200 mM NaCl) supplemented with freshly added proteinase K (Thermo Scientific) at a final concentration of 100 ng/ml. The cell lysates were extracted twice with a buffer-saturated phenol:chloroform:isoamyl alcohol mixture (25:24:1) and once with chloroform. Next, the genomic DNA was precipitated by the addition of 2.5 volumes of absolute ethanol and 0.5 volumes of 7.5 M ammonium acetate (pH 5.5). After washing with 70% ethanol, the DNA pellets were air dried and dissolved in 100 μl of Tris-EDTA buffer (10 mM Tris, pH 8.0, and 1 mM EDTA) supplemented with RNase A (Thermo Scientific) at a final concentration of 100 μg/ml. The genomic DNA of EGFP$^+$ clones derived from cultures transfected with plasmid DNA as well as that derived from cultures cotransduced with AdV.Cas9, AdV.gRNA[S1] and AdV.Δ2.donor[S1] was extracted by using the DNeasy Blood & Tissue Kit according to the

6

protocol provided by the manufacturer. The genomic DNA samples of FP635+ H27 clones were also isolated by deploying the DNeasy Blood & Tissue Kit.

PCR analyses of gene-targeting experiments with plasmid, IDLV and AdV donor DNA. The composition of the PCR mixtures and cycling parameters used for the analyses of genome-modification events are specified in the **Supplementary Tables 1** and **2**, respectively.

Statistics. Data sets were analyzed by using the GraphPad Prism 5 software package and evaluated for significance by applying unpaired two-tailed Student's *t*-tests (*P* < 0.05 considered significant).

Southern blot analyses. Genomic DNA was isolated from individual human myoblast clones and from parental human myoblast populations according to aforementioned organic solvent-based protocol. Next, DNA samples (10 µg or 20 µg) were digested overnight with *Nco*I (Thermo Scientific) and were resolved through a 0.8% agarose gel in 1× Tris-acetate-EDTA buffer. The DNA was transferred by capillary action onto an Amersham Hybond-XL membrane (GE Healthcare Life Sciences) using standard Southern blot techniques. The 635- and 790-bp DNA probes specific for the centromeric *AAVS1* arm and for the transgene in donor DNA templates, respectively, were isolated from source plasmids following standard restriction enzymes digestions and agarose gel electrophoreses. The purified DNA probes were subsequently radiolabeled with [α-^{32}P]dATP (GE Healthcare Life Sciences) by using the DecaLabel DNA labeling Kit following the manufacturer's instructions (Thermo Scientific). Prior to its deployment, the radiolabeled probe were separated from unincorporated dNTPs through size-exclusion chromatography in Sephadex-50 columns (GE Healthcare Life Sciences). A Storm 820 Phosphoimager (Amersham Biosciences) was used for the detection of the probe-hybridized DNA. The images were acquired by using the Storm Scanner Control 5.03 software and were processed with the aid of ImageQuant Tools 3.0 software (both from Amersham Biosciences).

COBRA-FISH karyotyping. The COBRA-FISH karyotyping of HeLa target cells was done according to a published protocol[55].

Analysis of *in vivo* excision of AdV donor DNA. The experiments designed to assess TALEN-mediated excision of donor[S1] DNA from AdV backbones were performed as follows. Eighty-thousand HeLa cells were seeded in wells of 24-well plates, and the following day they were cotransduced with AdV.Δ2.TALEN-L[S1] (1.5 TU/cell), AdV.Δ2.TALEN-R[S1] (1.5 TU/cell) and AdV.Δ2.donor[S1/T-TS] (3 TU/cell) or were cotransduced with 2 gene switch–activating units per cell of hcAd.FLPe.F[50] (ref. 56) and AdV.Δ2.donor[S1/FRT] (3 TU/cell). Control samples were provided by parallel HeLa cell cultures exposed exclusively to AdV.Δ2.donor[S1/T-TS] (3 TU/cell) or to AdV.Δ2.donor[S1/FRT] (3 TU/cell). At 72 h post transduction, extrachromosomal DNA was isolated essentially as described previously[57], after which 2-µl DNA samples were subjected to PCR. The PCR mixtures consisted of 0.4 µM of primer #997 (5′-GCACTGAAACCCTCAGTCCTAGG-3′), 0.4 µM of primer #998 (5′-CGGCGTTGGTGGAGTCC-3′), 0.1 mM of each dNTP (Invitrogen),

1 mM MgCl$_2$ (Promega), 1× Colorless GoTaq Flexi Buffer (Promega) and 2.5 U of GoTaq Flexi DNA polymerase (Promega). Next, 50-µl PCR mixtures were subjected to an initial 2-min denaturation period at 95 °C, followed by 30 cycles of 30 s at 95 °C, 30 s at 62 °C and 45 s at 72 °C. The reactions were terminated by a final extension period of 5 min at 72 °C. The detection of the resulting PCR products was performed by conventional agarose gel electrophoresis.

Investigation of vector-vector illegitimate recombination in a myoblast-to-myotube cellular differentiation model. Two-hundred thousand human myoblasts were seeded in wells of 24-well plates. Two days later, the cells were transduced with 10 TU/cell of AdV. Δ2.donor[S1] or with 10 TU/cell of IDLV.donor[S1] mixed with AdV.TALEN-L[S1] plus AdV.TALEN-R[S1] at MOIs of 2.5 TU/cell each or mixed with AdV.TALEN-R[S1] alone at an MOI of 5 TU/cell. Mock-transduced myoblasts served as negative controls. At 3 d post transduction, the growth medium[44] was removed, the cells were washed twice with 2 ml of PBS, and mitogen-free differentiation medium[44] was added onto the different myoblast cultures. Five days later, cellular DNA was extracted from the various cultures containing post-mitotic syncytial myotubes. Next, PCR analysis of the cellular DNA was carried out for detecting head-to-tail vector DNA forms using as negative controls nuclease-free water and DNA extracted from the mock-transduced myotube cultures. Positive controls were obtained by using *in vitro*–assembled head-to-tail AdV DNA templates (**Fig. 3c** and **Supplementary Fig. 4b**) and DNA isolated from IDLV.donor[S1]-transduced myoblast clone 10. The integrity of the various DNA templates was controlled for by parallel *DMD*-specific PCR amplifications.

Accession codes. GenBank/EMBL/DDBJ: AP45_pLV.donorEGFP, KF419293; AQ25_pLV.donorS1, KF419294. Plasmids available at Addgene: hCas9, 41815; gRNA_AAVS1-T2, 41818.

6

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

M.H. and I.M. contributed equally to this work. M.H. and I.M. generated reagents and performed most of the experiments with the help of S.F.D.H., J.M.J. and M.A.F.V.G.. T.C. generated and validated *AAVS1*-specific TALENs and *EGFP*-specific ZFNs. M.H., I.M., J.M.J. and M.A.F.V.G. designed the experiments and analyzed the data. M.A.F.V.G. conceived and initiated the research. M.A.F.V.G. wrote the manuscript with the help from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests.

REFERENCES

1. Capecchi, M.R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat. Rev. Genet*. **6**, 507-512 (2005).

2. Itzhaki, J.E. & Porter, A.C. Targeted disruption of a human interferon-inducible gene detected by secretion of human growth hormone. *Nucleic Acids Res*. **19**, 3835-3842 (1991).

3. Brown, J.P., Wei, W. & Sedivy, J.M. Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. *Science* **277**, 831-834 (1997).

4. Russell, D.W. & Hirata, R.K. Human gene targeting by viral vectors. *Nat. Genet*. **18**, 325-330 (1998).

5. Segal, D.J. & Meckler, J.F. Genome Engineering at the Dawn of the Golden Age. *Annu. Rev. Genomics Hum. Genet*. **14**, 135-158 (2013).

6. Biasco, L., Baricordi, C. & Aiuti, A. Retroviral integrations in gene therapy trials. *Mol. Ther*. **20**, 709-716 (2012).

7. Varga, C.M. *et al*. Quantitative comparison of polyethylenimine formulations and adenoviral vectors in terms of intracellular gene delivery processes. *Gene Ther.* **12**, 1023-1032 (2005).

8. Wanisch, K. & Yáñez-Muñoz, R.J. Integration-deficient lentiviral vectors: a slow coming of age. *Mol. Ther*. **17**, 1316-1332 (2009).

9. Cornu, T.I. & Cathomen, T. Targeted genome modifications using integrase-deficient lentiviral vectors. *Mol. Ther*. **15**, 2107-2113 (2007).

10. Lombardo, A. *et al*. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol*. **25**, 1298-1306 (2007).

11. Gabriel, R. *et al*. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol*. **29**, 816-823 (2011).

12. Lombardo, A. *et al*. Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods* **8**, 861-869 (2011).

13. Benabdallah, B.F. *et al*. Targeted gene addition of microdystrophin in mice skeletal muscle via human myoblast transplantation. *Mol. Ther. Nucleic Acids* **2**, e68 (2013).

14. Coluccio, A. *et al*. Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther*. **21**, 1695-1704 (2013).

15. Mussolino, C. *et al*. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res*. **42**, 6762-6773 (2014).

16. Holkers, H. *et al*. Differential integrity of *TALE* nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. **41**, e63 (2013).

17. Nightingale, S.J. *et al*. Transient gene expression by nonintegrating lentiviral vectors. *Mol. Ther*. **13**, 1121-1132 (2006).

18. Mátrai, J. *et al*. Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. *Hepatology* **53**, 1696-707 (2011).

19. 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).

20. Li, H. *et al*. *In vivo* genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217-221 (2011).

21. Ellis, B.L., Hirsch, M.L., Porter, S.N., Samulski, R.J. & Porteus, M.H. Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by Food and Drug Administration-approved drugs. *Gene Ther*. **20**, 35-42 (2013).

22. Rahman, S.H. *et al*. The nontoxic cell cycle modulator indirubin augments transduction of adeno-associated viral vectors and zinc-finger nuclease-mediated gene targeting. *Hum. Gene Ther*. **24**, 67-77 (2013).

6

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

24. Gonçalves, M.A.F.V. & de Vries, A.A.F. Adenovirus: from foe to friend. *Rev. Med. Virol*. **16**, 167-186 (2006).

25. Terns, R.M. & Terns, M.P. CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends Genet*. **30**, 111-118 (2014).

26. Fu, Y. *et al*. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol*. **31**, 822-826 (2013).

27. Hsu, P.D. *et al*. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol*. **31**, 827-832 (2013).

28. Cradick, T.J., Fine, E.J., Antico, C.J. & Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res*. **41**, 9584-9592 (2013).

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

30. Chadeuf, G., Ciron, C., Moullier, P. & Salvetti, A. Evidence for encapsidation of prokaryotic sequences during recombinant adeno-associated virus production and their in vivo persistence after vector delivery. *Mol. Ther*. **12**, 744-753 (2005).

31. Miller, J.C. *et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol*. **25**, 778-85 (2007).

32. Szczepek, M *et al*. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol*. **25**, 786-793 (2007).

33. Doyon, Y. *et al*. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* **8**, 74-79 (2011).

34. Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **8**, 765-770 (2011).

35. Petek, L.M., Russell, D.W. & Miller, D.G. Frequent endonuclease cleavage at off-target locations *in vivo*. *Mol. Ther*. **18**, 983-986 (2010).

36. Miller, D.G. *et al*. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J. Virol*. **79**, 11434-11442 (2005).

37. Nakai, H. *et al*. Helper-independent and AAV-ITR-independent chromosomal integration of double-stranded linear DNA vectors in mice. *Mol. Ther*. **7**, 101-111 (2003).

38. Choi, V.W., McCarty, D.M. & Samulski, R.J. Host cell DNA repair pathways in adeno-associated viral genome processing. *J. Virol*. **80**, 10346-10356 (2006).

39. Duffy, S., Shackelton, L.A. & Holmes, E.C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet*. **9**, 267-276 (2008).

40. Khan, I.F., Hirata, R.K. & Russell, D.W. AAV-mediated gene targeting methods for human cells. *Nat. Protoc*. **6**, 482-501 (2011).

METHODS REFERENCES

41. Gonçalves, M.A.F.V., van der Velde, I., Knaän-Shanzer, S., Valerio, D. & de Vries, A.A.F. Stable transduction of large DNA by high-capacity adeno-associated virus/adenovirus hybrid vectors. *Virology* **321**, 287-296 (2004).

42. Fallaux, F.J. *et al*. New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Hum. Gene Ther*. **9**, 1909-1917 (1998).

43. Havenga, M.J. *et al*. Serum-free transient protein production system based on adenoviral vector and PER.C6 technology: high yield and preserved bioactivity. *Biotechnol. Bioeng*. **100**, 273-283 (2008).

44. Cudré-Mauroux, C. *et al*. Lentivector-mediated transfer of Bmi-1 and telomerase in muscle satellite cells yields a duchenne myoblast cell line with long-term genotypic and phenotypic stability. *Hum. Gene Ther*. **14**, 1525-1533 (2003).

45. Gonçalves, M.A.F.V. *et al*. Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. *Mol. Ther*. **19**, 1331-1341 (2011).

46. Coluccio, A. *et al*. Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther*. **21**, 1695-1704 (2013).

47. Janssen, J.M., Liu, J., Skokan, J., Gonçalves, M.A.F.V. & de Vries, A.A.F. Development of an AdEasy-based system to produce first- and second-generation adenoviral vectors with tropism for CAR- or CD46-positive cells. *J. Gene Med*. **15**, 1-11 (2013).

48. Maggio, I., Holkers, M., Liu, J., Janssen, J.M., Chen, X. & Gonçalves M.A.F.V. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. *Sci. Rep*. **4**, 5105 (2004).

49. Mali, P. *et al*. RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

50. Holkers, H. *et al*. Differential integrity of *TALE* nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. **41**, e63 (2013).

51. Holkers, M., Cathomen, T. & Gonçalves M.A.F.V. Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells. *Methods*, DOI 10.1016/j.ymeth.2014.02.017.

52. Pelascini, L.P.L. *et al*. Histone deacetylase inhibition rescues gene knockout levels achieved with integrase-defective lentiviral vectors encoding zinc-finger nucleases. *Hum. Gene Ther. Methods* **24**, 399-411 (2013).

53. Pelascini, L.P.L., Janssen, J.M. & Gonçalves M.A.F.V. Histone deacetylase inhibition activates transgene expression from integration-defective lentiviral vectors in dividing and non-dividing cells. *Hum. Gene Ther*. **24**, 78-96 (2013).

54. Pelascini, L.P.L. & Gonçalves M.A.F.V. Lentiviral Vectors Encoding Zinc-Finger Nucleases Specific for the Model Target Locus *HPRT1*. *Methods Mol. Biol*. **1114**, 181-199 (2014).

55. Brielmeier, M. *et al*. Improving stable transfection efficiency: antioxidants dramatically improve the outgrowth of clones under dominant marker selection. *Nucleic Acids Res*. **26**, 2082-2085 (1998).

56. van Nierop, G.P., de Vries, A.A.F., Holkers, M., Vrijsen, K.R & Gonçalves, M.A.F.V. Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Res*. **37**, 5725-5736 (2009).

57. Szuhai, K. & Tanke HJ. COBRA: combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. *Nat. Protoc*. **1**, 264-275 (2006).

58. Gonçalves, M.A.F.V. *et al*. Targeted chromosomal insertion of large DNA into the human genome by a fiber-modified high-capacity adenovirus-based vector system. *PLoS ONE* **3**, e3084 (2008).

59. Holkers, M, de Vries, A.A.F. & Gonçalves, M.A.F.V. Nonspaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells. *Nucleic Acids Res*. **40**, 1984-1999 (2012).

6

7

Summary
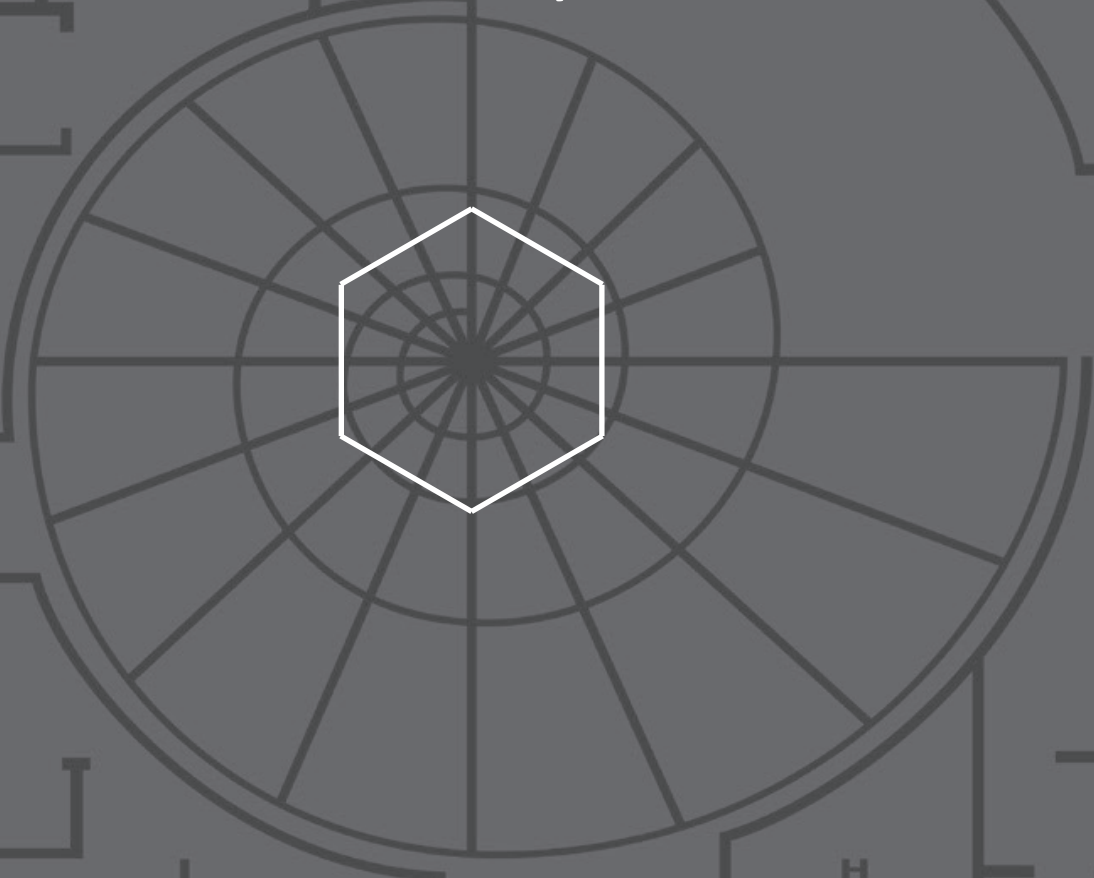
INTRODUCTION

Providing a cure for monogenetic diseases, not by means of administering protein replacement therapies, but by gene complementation or even gene repair instead, is a long sought goal of gene therapy. Some of these expectations were met for specific genetic disorders by providing corrective expression units into hematopoietic stem cells which, once transduced *ex vivo* and re-infused into the patient, could repopulate the bone marrow providing the blood system with corrected cells [1-3]. The success of these trials were clouded with the onset of leukemia in some patients initiated by the insertion of the viral vectors used pushing scientists back to the drawing board in order to redesign their genetic carriers [4]. After many years of optimization, involving trial and error experimentation and extensive laboratory work, has led to, nowadays, vector systems with an highly improved safety profile. These developments combined with extensive screening of patients enrolled in clinical trials, is opening the perspective for tackling many genetic diseases through gene therapy. [5, 6]. Despite this trust in the field of gene therapy, the research is set to continue gathering momentum. For the time being, the gene complementation strategy on itself gives clinicians a tool to tackle genetic defects, but the ultimate goal of gene therapy is to directly repair the genetic defect by changing the endogenous mutated DNA into a functional gene. Currently, very large efforts are ongoing to achieve this DNA repair by activating the cellular DNA damage response pathways following the generation of sequence-specific lesions in the chromosomal DNA of human cells. These targeted DNA breaks can subsequently processed by the non-homologous end-joining DNA repair pathway resulting in small insertions and deletions (indels). Alternatively, in the presence of exogenous donor DNA templates, the targeted DNA breaks can be processed by the homologous recombination (HR) pathway and can lead to precise chromosomal DNA changes by the removal of a disease-causing mutation [7, 8].

Of note, these genome manipulation processes can be initiated by the formation of a single- or double-stranded DNA break in the target region of choice [9 - 11]. Work done during the last two decades of the previous century showed that by inducing a double-stranded DNA break, the restoration of an expression unit by homology-directed gene repair (or HR) could be enhanced from about 10^{-6} events to up to 10% of the targeted population. Designer nucleases like zinc-finger nucleases (ZFNs) [12], transcription activator-like effector nucleases (TALENs) [13] and clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins (Cas) [14-16] are capable of generating these site-specific DNA lesions. These various technologies have their own sets of pros and cons in terms of specificity, complexity and the amenability for multiplexing, yet the general principle is to transiently express an exogenous enzyme in target cells. A potential alternative research line

to the use of designer nucleases, is to investigate whether specific DNA structures can, by themselves, serve as triggers of the DNA damage response and, in doing so, elicit targeted gene repair. Such an approach would simplify genome editing protocols, such as, by reducing the number of reagents needed to be introduced into target cells. The discovery of GEN1 and SLX4 complexes involved in the resolution of Holliday-like secondary DNA structures [17, 18], led us to investigating whether such high-order DNA structures can serve as stimuli for HR. Specifically, we hypothesized that recognition and processing of secondary DNA forms associated with an expression unit by cellular gene products like the aforementioned GEN1 and SLX4 would lead to their resolution by DNA cleaving and ensuing targeted gene repair.

In **chapter 2** of this thesis, evidence is provided for the capacity of relatively small, non-spaced, inverted repeats located within an open reading frame to provoke HR-mediated DNA repair. By using a complementary donor-template, restoration of the coding region could be demonstrated by measuring transgene expression and DNA sequence analysis. In this chapter, an extrachromosomal functional read-out system, based on pairs of complementary DNA templates carrying defective reporter gene sequences, was devised, that can serve as substrates for testing intermolecular HR-dependent gene repair. This experimental system allowed us to investigate in a quantitative manner the effect of various types of single DNA repeats on the HR process in mammalian cells as well as the role of DNA replication on the recombinogenic potential of these motifs. We found that, in contrast to direct and spaced inverted repeats, both simple palindromes and composite inverted DNA repeats constitute targets for the HR pathway in mammalian cells. Induction of homology-directed gene repair was dependent on the arrangement and spacing of the repetitive DNA unit rather than on its nucleotide sequence. We also found that the presence of inverted DNA repeat sequences in target molecules rendered them susceptible to coordinated nicking by T7 endonuclease I, a *bona fide* four-way DNA branch resolving enzyme [19]. These results are consistent with other *in vitro* data showing that lineform-to-cruciform transition in double-stranded DNA molecules relies on the presence of an inverted repeat and is negatively affected by intervening spacer sequences in a length-dependent manner [20-22]. Furthermore, as aforementioned, by endowing acceptor DNA molecules with an eukaryotic origin of replication, we could probe in a strict manner the role of template DNA synthesis on repeat-induced homology-directed gene repair. Thus, in this chapter, it was demonstrated that non-spaced inverted DNA repeats can *per se* stimulate homology-directed gene repair in mammalian cells presumably due to their capacity to form secondary structures *in vivo* that can subsequently serve as direct targets for cellular structure-specific nucleases.

These findings form a basis to further investigate the usefulness of secondary DNA structures in gene editing protocols. However, the presence of these secondary DNA structures in target

loci is scarce and reduces the use of this technique for the engagement of the DNA-repair pathways. Being able to induce the DNA repair in a site-specific manner As mentioned earlier, the efficient HR-mediated chromosomal insertion of donor DNA templates is dependent on inducing DNA lesions at specific target sequences. To this end, designer nucleases like ZFNs, TALENs and CRISPR/Cas-based nucleases must be introduced into target cells. Therefore, in this Thesis, different methodologies of providing target cells with expression units encoding these molecular scissors, were investigated. The starting of this research, described in **chapter 3**, consisted of investigating the feasibility in producing lentiviral vectors (LVs) and adenoviral vectors (AdVs) encoding TALENs. It was found that HIV-1-based LV genomes bearing *TALEN* sequences are predisposed to rearrangements in target cells. The structural analyses of these genomes in HeLa cell populations and in individual HeLa cell clones by PCR, Southern blotting and DNA sequencing showed that most of the rearrangements occurred through recombination events involving the *TALE* repeat array, ultimately leading to deletions with various sizes. The DNA sequence analyses in particular revealed that these rearrangements of the *TALE* array, albeit variable in number, consisted of precise as opposed to randomly truncated deletions of individual *TALE* repeats. Presumably, these deletions/truncations occur during the conversion of the lentiviral vector RNA genomes into complementary DNA by the viral reverse transcriptase (RT). The low template processivity of RT, combined with its low template affinity caused by RNaseH-mediated removal of the RNA template, leads to template switching events and ensuing deletions of *TALE* repeat sequences. Indeed, subsequent independent experiments performed by others provided additional evidence for this interpretation which can be reduced by different means. It has been recently demonstrated that the genetic instability of *TALEN* ORFs following lentiviral vector transductions can be overcome, either by mutating the RT, effectively converting the lentiviral RNA genomes in mRNA templates for protein production [23], or by re-coding the *TALE* repeats in order to minimizing as much as possible identical sequences and thereby reduce their "repetitiveness" character [24]. Conversely, data presented also in Chapter 3, demonstrates that *TALEN* sequences are stably maintained in first- and second-generation adenoviral vectors following their serial propagation in producer cells. Indeed, the TALEN-encoding transgenes present in the adenoviral vector genomes suffered neither large rearrangements nor small-scale mutations as revealed by restriction fragment length analysis and DNA sequencing, respectively. Importantly, transduction experiments with the resulting TALEN-encoding adenoviral vector preparations led to DSB formation at the intended target chromosomal locus in transformed and non-transformed human cells at similarly high frequencies. A detailed protocol for the introduction of *TALEN* expression units into AdV backbones, and for the production, purification and characterization of the corresponding vector preparations is described in **chapter 4**.

7

In **Chapter 5**, the feasibility of using AdVs for the delivery of CRISPR/Cas9 components into human cells is investigated. To this end, AdV preparations containing either an expression unit for the *S. pyogenes* Cas9 nuclease or for an RNA Pol-III polymerase expression unit of an *AAVS1*-specific or *eGFP*-specific guide RNA, required to direct the Cas9 nuclease complex to its complementary DNA target sequence were produced and characterized. The capability for AdVs encoding RGN components, to achieve robust targeted mutagenesis in a diverse array of human cell types was clearly demonstrated. In all the cell types tested, the dose-dependent CRISPR/Cas9 mediated disruption of the target sequence was demonstrated. In specific examples, there was a similar efficiency measured when compared to an, also AdV delivered, isogenic *AAVS1*-specific TALEN pair. In contrast, through co-transfection experiments in human pluripotent stem cell lines comparing gRNA:Cas9 sets with TALEN pairs based on specific architectures, Ding and colleagues showed that, for each targeted locus, the RGNs consistently and substantially outperformed the TALENs [25]. Nonetheless, these results do support the view that the TALEN platform is not per se inferior to that of RGNs in what site-specific chromosomal cleaving activity is concerned.

Taken together, the data presented in **Chapters 3**, **4** and **5** provide a concrete basis for the use of AdV-based technology for the delivery of artificial nucleases into human target cells. In gene editing protocols depending on NHEJ events, the introduction of these nucleases is enough. Nevertheless, for the repair of genes, the introduction of an exogenous DNA template provided at the same time as the nucleases is needed. In **chapter 6**, the impact of different types of donor DNA templates on the efficiency, specificity and accuracy of the genome editing process was determined. Next to the efficiency of donor DNA template usage, the accuracy and specificity can have a great impact on the outcome of the repaired sequences on an individual cell basis. For this assessment, a donor template, consisting of an eGFP expression cassette, was used. This unit was flanked by sequences homologous to the human *AAVS1* locus in which the target site of TALENs or CRISPR/Cas9 was situated. Transducing DMD myoblasts with IDLV particles, carrying the *AAVS1*-targeting donor DNA, combined with AdV-mediated delivery of the AAVS1-specific TALEN pair, yielded a relatively large population of cells (i.e. 9%) stably expressing the transgene. This stably transduced population contained, however, several cells fractions whose donor DNA had inserted into the genome via "illegitimate", recombination events. Moreover, although the majority of stably transduced cells contained exogenous DNA insertions formed by HR at termini precise homologous recombination (76,0%), 33,7% of this population contained two or more head-to-tail concatemeric copies of the donor DNA. These concatemeric forms are likely to contribute to heterogeneous expression levels in stably transduced cell populations and, especially in a gene repair settings, are expected to be deleterious for the success of the gene editing approach. When the target cells were subjected to the same set of

AAVS1-specific TALENS but were instead transduced with an AdV delivering the *AAVS1*-targeting donor DNA template, the percentage of stably transduced cells was lower that obtained following the use of IDLV donor DNA (i.e. 1,24% versus 9%). Nevertheless, detailed molecular analysis of individual single cell-derived clones revealed that AdV-modified cells underwent through a precise genome editing process. Specifically, in all randomly selected clones analyzed, the donor DNA templates were found to be properly inserted at the target site, that is, endogenous-exogenous DNA junctions were formed by HR events and there without evidence for concatemerization. Furthermore, different topologies of plasmid DNA were also used as source for donor-DNA templates. Finally, the designer nuclease-assisted genome editing specificity and accuracy attained by using plasmid donor DNA templates was also investigated. It was found that the topology of the plasmid template has a clear impact on the specificity and accuracy of the genome editing process. Indeed, supercoiled plasmids were less prone to non-homologous recombination events when compared with linearized free-ended plasmid templates. In any case, neither supercoiled nor linear donor DNA plasmids yielded the levels of genome editing specificity and accuracy attainable by using AdV donor DNA templates. The findings summarized in this thesis put forward the view that the numerous efforts devoted to minimizing off-target activity of sequence-specific nucleases, should be complemented with those aiming at identifying HR substrates whose features maximize on-target and accurate insertion of foreign DNA. The development of these optimized HR templates is expected to promote fundamental and applied research activities dependent on the precise manipulation of mammalian genomes.

FUTURE DIRECTIONS

In the future, patients suffering from a genetic disorder will go to their medical specialists, which can provide them with a patient-specific treatment designed to repair the mutated sequences in the affected cell population. This is, however, until now only science fiction. Decades of research in the field of gene therapy did bring this "fiction" closer to reality by being able to complement mutated sequences with correct genes packaged into integrating viral vector particles. Upon transduction of target cells, these genomes can integrate and express the wild-type protein, effectively curing the genetic disease. This large first step in gene therapy buys time to optimize the ultimate goal, being able to precisely repair the mutation, restoring the endogenous gene and allowing cellular control over the expression. How far are we from reaching this goal? In a laboratory setting, using specialized cell types and molecular techniques, researches are able to effectively transduce these cells with nucleases and donor-DNA templates resulting in gene repair in a rather small target cell population fraction of clinically relevant cells. These levels are, until now, not high enough

to allow translation into most clinical trial protocols. All of these elements on themselves have been proven to be efficient; nuclease systems such as ZFNs, TALENs or CRISPR/Cas-based nucleases can all disrupt their target sites with very high frequency and donor-DNA templates can be adopted by the cells in frequencies close to 10% in immortalized cells. Nevertheless, bringing all these components together is one of the, if not the biggest, hurdle to take to bring gene repair into the clinic.

The unsurpassed efficiency of viral vectors in transducing target cells, introducing the individual components for gene repair, makes them prime candidates to take this hurdle. In this thesis, the use of AdVs for the delivery of each of the modules needed for gene editing, i.e., the artificial nucleases and the donor-DNA templates, is demonstrated. In the future, research will focus on being able to package all of these components into a single viral vector particle. Furthermore, as the problem of efficient delivery starts to be satisfactorily tackled, enhancing the efficiency of homology-directed gene repair itself is required in order to reach clinical relevant levels. For instance, by blocking the error-prone non-homologous end-joining pathway to maximize HR-mediated repair of site-specific DSBs, might increase the rate of gene corrected cells. Nevertheless, as these DNA repair pathways share a number of proteins, the challenge will be to determine which signals to block in order to let the HR pathway to prevail. In a recent publication, it was shown that knocking down proteins involved in NHEJ, the rate of CRISPR mediated HR was elevated 4 to 5-fold [26]. Interestingly, by expressing the adenoviral 4 proteins E1B55K and E4orf6 improved the efficiency up to eightfold in human and mouse cell lines. However, being able to specifically transduce, *in vitro*, but even more challenging *in vivo*, clinically relevant cell types in an highly efficient manner, followed by proficient engagement of the DNA repair machinery, will take considerable amount of research. Yet, as the rapid progression in the search of efficient gene editing protocols demonstrates, the next step in gene therapy, gene repair, is merely a matter of time.

REFERENCES

1. Cavazzana-Calvo, M. *et al*. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science. 288, 669-672 (2000)

2. Hacein-Bey, S., Yates, F., de Villartay, J.P., Fischer, A. & Cavazzana-Calvo, M. Gene therapy of severe combined immunodeficiencies: from mice to humans. Neth. J. Med. 60, 299-301 (2002)

3. Aiuti, A *et al*. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. Science. 296, 2410-2413 (2002)

4. Hacein-Bey-Abina, S. *et al*. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. 118, 3132-3142 (2008)

5. Aiuti, A. *et al*. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. Science. 341, 1233151 (2013)

6. Biffi, A. *et al*. Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science. 341, 1233158 (2013)

7. Genovese, P. *et al*. Targeted genome editing in human repopulating haematopoietic stem cells. Nature 510, 235-240 (2014)

8. Perez, E.E. *et al*. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat. Biotechnol. 26, 808-816. (2008)

9. Jasin, M. & Berg, P. Homologous integration in mammalian cells without target gene selection. Genes Dev. 2, 1353-1363. (1988)

10. Lukacsovich T, Yang D, Waldman AS. Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. Nucleic Acids Res. 1994;22:5649-5657

11. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell Biol. 14, 8096-8106. (1994)

12. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci USA. 93, 1156-60 (1996)

13. Christian, M. *et al*. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics. 186, 757-761 (2010)

14. Cong, L. *et al*. Multiplex genome engineering using CRISPR/Cas systems. Science. 339, 819-823 (2013)

15. Mali, P. *et al*. RNA-guided human genome engineering via Cas9. Science. 339, 823-826 (2013)

16. Jinek, M. *et al*. RNA-programmed genome editing in human cells. Elife. 29, e00471. (2013)

17. Ip, S.C. *et al*. Identification of Holliday junction resolvases from humans and yeast. Nature. 456, 357-361. (2008)

18. Svendsen, J.M. *et al*. Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell. 138, 63-77. (2009)

19. Declais, A.C. and Lilley, D.M.J. New insight into the recognition of branched DNA structure by junction-resolving enzymes. Curr. Opi. Struct. Biol. 18, 86–95. (2008)

20. Vologodskii, A.V. and Kamenetskii, F. The relaxation time for a cruciform structure in superhelical DNA. FEBS Lett. 160, 173–176. (1983)

21. Sinden, R.R., Zheng, G.X., Brankamp, R.G. and Allen, K.N. On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. Genetics 129, 991–1005.(1991)

22. Kogo, H. *et al*. Cruciform extrusion propensity of human translocation-mediating palindromic AT-rich repeats. Nucleic Acids Res., 35, 1198–1208. (2007)

23. Mock, U. *et al*. Novel lentiviral vectors with mutated reverse transcriptase for mRNA delivery of TALE nucleases. Sci. Rep. 4, 6409. (2014)

7

24. Yang, L. *et al*. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 41, 9049-9061. (2013)

25. Ding, Q. *et al*. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell 12, 393–394 (2013).

26. Chu, V.T. *et al*. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. Mar 24 (2015)

Addendum

Nederlandse samenvatting
List of publications
Curriculum Vitae
Dankwoord

NEDERLANDSE SAMENVATTING

Sommige genetische aandoeningen, zoals bijvoorbeeld de ziekte van Pompe, kunnen bestreden worden door het aanbieden van medicijnen die de taak van het defecte of ontbrekende eiwit overnemen. Deze methode is kostbaar en belastend voor de patiënt. Gentherapie is een andere methode om dit soort genetische aandoeningen te behandelen. Hierbij wordt een correctie kopie van het defecte gen in de cel gebracht of wordt het defecte gen zelf gerepareerd in de aangedane cellen. In de jaren '90 van de vorige eeuw hebben onderzoekers het principe van gen-complementatie succesvol toegepast bij patiënten die lijden aan een aandoening van het immuunsysteem. Een correcte kopie van het gen werd met behulp van een virale vector in bloedcellen gebracht waar het, na integratie, kon zorgen voor de aanmaak van functioneel eiwit.

Het succes van deze klinische studies werd helaas overschaduwd door de ontwikkeling van leukemie in enkele van de patiënten. Na onderzoek bleek integratie van de virale vector op een verkeerde plaats de oorzaak van de ontsporing van de witte bloedcellen. Nieuwe ontwikkelingen op het gebied van vector-ontwerp en diepgaande moleculaire screening van de gemodificeerde cellen heeft ervoor gezorgd dat er hedendaags vectoren beschikbaar zijn met een verhoogde veiligheidsprofiel waardoor deze vorm van patiënt-specifieke medische zorg steeds vaker toegepast kan worden met acceptabele risico's voor de patiënt.

Naast deze vorm van gentherapie, de gen-complementatie, is het mogelijk om de sequentie van genen zodanig aan te passen dat mutaties gerepareerd kunnen worden. Deze vorm van gen-reparatie is complex en afhankelijk van de activatie van cellulaire eiwitten betrokken bij de reparatie van DNA schade. Door deze activatie kan de machinerie exogene donor-DNA fragmenten gebruiken om het genoom te repareren. De kans op spontane activatie van deze cascade op de exacte locatie van het gendefect is vrijwel nihil. Door onderzoek in gistcellen is duidelijk geworden dat deze activatie kan worden getriggerd door het aanbrengen van een enkel- of dubbelstrengs breuk in het DNA. Middelen om deze breuk te genereren zijn tegenwoordig voorhanden in de vorm van artificiële nucleasen waarbij DNA digestie domeinen van restrictie enzymen naar de locatie van keuze worden gebracht door zink-vingers (ZFNs) of zogenoemde "transcription activator-like effector nucleases", TALENs. De nieuwste toevoeging aan dit pallet is het CRISPR/Cas systeem, wat staat voor "clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins (Cas)". Deze nucleasen worden doormiddel van korte stukken RNA, "guide RNA" genoemd, naar de exacte locatie gebracht. In **hoofdstuk 2** van deze thesis vroegen we ons af of deze breuken ook gegenereerd konden worden doormiddel van DNA sequenties zelf in plaats van het aanmaken van niet-humane eiwitten. Op basis van pas gepubliceerde resultaten waarbij onderzoekers lieten zien dat een eiwit genaamd GEN1 specifieke secundaire DNA structuren kon herkennen en ontvouwen, construeerden we twee plasmiden, een

acceptor en een donor. De acceptor plasmide bevat een relatief korte haarspeldstructuur in een hrGFP expressie cassette. Pas na recombinatie met een andere plasmide die de homologe sequenties bevatte zonder deze haarspeldstructuur, is het hrGFP gen functioneel en kleuren getransduceerde cellen fluorescent groen. In **hoofdstuk 2** toont aan dat deze haarspeldstructuren homologe recombinatie activeren tussen deze twee plasmiden en dat dezelfde sequentie, gekloneerd in een directe oriëntatie of een haarspeldstructuur met een DNA fragment ertussen, deze recombinatie niet activeerden. Uitgebreide moleculaire karakterisering van de gerecombineerde plasmiden alsmede de bepaling van het mechanisme toonden aan dat deze recombinatie sequentie-onafhankelijk was, en onafhankelijk van replicatie van de plasmiden, maar afhankelijk van de haarspeldstructuur. Hoewel deze resultaten aangaven dat haarspeldstructuren in het DNA gebruikt kunnen worden om exogene DNA fragmenten in te brengen bleek na literatuurstudie dat deze structuren zeldzaam zijn. Vrijwel alle recombinogene sequenties zijn, tijdens de evolutie, uit het genoom verdwenen.

In **hoofdstuk 3** geeft weer dat de keuze voor een type virale vector een grote invloed kan hebben op de stabiliteit van TALEN gen sequenties na transductie in doelcellen. In deze studie zijn twee expressie cassettes voor *AAVS1*-specifieke TALENs in HIV-1 gebaseerde lentivirale vectoren (LV) geplaatst en in vectoren gebaseerd op adenovirussen (AdV). Waar de eerstgenoemde een grote diversiteit in ingekorte genomen laat zien, geven AdV getransduceerde cellen een hoge mate van TALEN activiteit weer. De hypothese was dat het kopiëren van het lentivirale RNA genoom in DNA, doormiddel van het eigen relatief slordige reverse transcriptase, deleties van specifieke stukken van het repetitieve deel van de TALEN met zich meebracht. Het sequensen van integreerde virale genomen toonde grote deleties, verschillend per kloon. Deze deleties waren heel specifiek en omvatten veelal een meervoud van 102 baseparen, waarbij elke 102 nucleotiden voor 1 variabel TALE domain codeerde. AdV getransduceerde cellen gaven dit patroon niet, integendeel, het genoom van de AdV was volledig intact en ook de eiwitten zelf waren functioneel in zowel getransformeerde als primaire cellen. **Hoofdstuk 4** beschrijft een uitvoerig protocol voor het kloneren van deze TALENs in AdV, en laat zien hoe de virale vectoren geproduceerd en gekarakteriseerd kunnen worden.

In **hoofdstuk 5** van deze thesis is dezelfde principe van het door AdV afleveren van nuclease componenten in doelcellen toegepast voor elementen van het CRISPR/Cas systeem. In afzonderlijke virale vectoren werden expressie cassettes voor Cas9 nuclease en de "guide RNA" geplaatst en gebruikt om doelcellen te transduceren. We hebben aangetoond dat de efficiëntie van CRISPR/Cas-gemedieerde digestie van de doelsequentie afhankelijk was van de dosis virale vector en dat per celtype de effectiviteit gelijk is aan of minder is dan TALEN-gemedieerde digestie van de doelsequentie. Dit staat in contrast met wat Ding en collega's publiceerden in "Cell Stem Cell" waarbij ze aantoonden dat doormiddel van transfectie in

humane pluripotente stamcellen, het CRISPR/Cas altijd superieur was aan TALENs. Onze mening is dat de efficiëntie van artificiële nucleasen gemedieerde digestie van sequenties in doelcellen sterk afhankelijk is van de wijze waarop de genen coderend voor deze eiwitten in doelcellen gebracht wordt.

In **hoofdstuk 6** van deze thesis hebben we de meest gangbare methodes voor het aanleveren van donor-DNA in doelcellen met elkaar vergeleken en daarbij gekeken naar de effectiviteit van gebruik, maar ook naar de moleculaire compositie van het eindresultaat. De totale populatie gemodificeerde cellen kan meerdere subpopulaties bevatten, elk met hun eigen sequentie, wat gevolgen kan hebben voor de uiteindelijke expressie van het gecorrigeerde gen. Als eerste werden de twee meest gebruikte virale vectoren getest, die van de integratie deficiëntie lentivirale vectoren (IDLV) en AdV. Hoewel de eerstgenoemde een veel hogere efficiëntie van transgen insertie liet zien (9,1% tegenover 1,24% in myoblasten) liet de moleculaire karakterisatie duidelijk zien dat slechts 42,3% van deze populatie een enkele, correct ingebrachte, kopie van een eGFP expressie cassette bevatte. Dit in tegenstelling tot de cellen getransduceerd met AdV deeltjes waarbij alle klonen deze juiste insertie liet zien. Pas nadat er genomisch instabiele HeLa cellen in combinatie met het intrinsiek slordige CRISPR/Cas nuclease systeem werd gebruikt konden er cellen gevonden worden die niet een juiste insertie bevatten. Als laatste werden ook plasmiden, in diverse conformaties, getest. Deze bleken dezelfde gemengde populatie te bevatten waarbij de circulaire vormen minder van deze onzorgvuldige inserties lieten zien.

Samengevat geven de **hoofdstukken 3**, **4**, **5** en **6** van deze thesis weer hoe AdVs gebruikt kunnen worden om individuele componenten van gentherapie efficiënt in doelcellen te brengen. Daarbij geeft het een basis om, naast de intensieve zoektocht naar steeds specifiekere nuclease systemen met gereduceerde "off-target" effecten, donor-DNA fragmenten te onderzoeken die bijdragen aan deze specificiteit. De resultaten uit **hoofdstuk 2** kunnen daaraan bijdragen; specifieke DNA samenstellingen dragen bij aan de activatie van reparatie van deze structuren wat mogelijk een toegevoegde waarde heeft om, wellicht, de efficiëntie van gen reparatie in de toekomst te verhogen.

S

LIST OF PUBLICATIONS

- **Adenoviral vector DNA for accurate genome editing with engineered nucleases**
  Holkers M*, Maggio I*, Henriques SF, Janssen JM, Cathomen T, Gonçalves MA. Nat. Methods. 2014

- **Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells.** Maggio I*, Holkers M*, Liu J, Janssen JM, Chen X, Gonçalves MA. Sci. Rep. 2014

- **Construction and characterization of adenoviral vectors for the delivery of TALENs into human cells.** Holkers M, Cathomen T, Gonçalves MA. Methods. 2014

- **Histone Deacetylase Inhibition Rescues Gene Knockout Levels Achieved with Integrase-defective Lentiviral Vectors Encoding Zinc-finger Nucleases.** Pelascini LP, Maggio I, Liu J, Holkers M, Cathomen T, Gonçalves MA. Hum. Gene Ther. Methods. 2013

- **Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells.** Holkers M, Maggio I, Liu J, Janssen JM, Miselli F, Mussolino C, Recchia A, Cathomen T, Gonçalves MA. Nucleic Acids Res. 2013

- **Concerted nicking of donor and chromosomal acceptor DNA promotes homology-directed gene targeting in human cells**. Gonçalves MA, van Nierop GP, Holkers M, de Vries AA. Nucleic Acids Res. 2012

- **Nonspaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells.** Holkers M, de Vries AA, Gonçalves MA. Nucleic Acids Res. 2012

- **Rapid and sensitive lentivirus vector-based conditional gene expression assay to monitor and quantify cell fusion activity.** Gonçalves MA, Janssen JM, Holkers M, de Vries AA. PLoS One. 2010

- **Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease.** van Nierop GP, de Vries AA, Holkers M, Vrijsen KR, Gonçalves MA. Nucleic Acids Res. 2009

- **Targeted chromosomal insertion of large DNA into the human genome by a fiber-modified high-capacity adenovirus-based vector system.** Gonçalves MA, <u>Holkers M</u>, van Nierop GP, Wieringa R, Pau MG, de Vries AA. PLoS One. 2008

- **Genetic complementation of human muscle cells via directed stem cell fusion**. Gonçalves MA, Swildens J, <u>Holkers M</u>, Narain A, van Nierop GP, van de Watering MJ, Knaän-Shanzer S, de Vries AA. Mol. Ther. 2008

- **Modular and excisable molecular switch for the induction of gene expression by the yeast FLP recombinase.** <u>Holkers M</u>, De Vries AA, Gonçalves MA. Biotechniques. 2006

- **Transduction of myogenic cells by retargeted dual high-capacity hybrid viral vectors: robust dystrophin synthesis in duchenne muscular dystrophy muscle cells.** Gonçalves MA, <u>Holkers M</u>, Cudré-Mauroux C, van Nierop GP, Knaän-Shanzer S, van der Velde I, Valerio D, de Vries AA. Mol. Ther. 2006

- **Human mesenchymal stem cells ectopically expressing full-length dystrophin can complement Duchenne muscular dystrophy myotubes by cell fusion.** Gonçalves MA, de Vries AA, <u>Holkers M</u>, van de Watering MJ, van der Velde I, van Nierop GP, Valerio D, Knaän-Shanzer S. Hum. Mol. Genet. 2006

CURRICULUM VITAE

De auteur van dit proefschrift werd op 10 mei 1980 geboren in Eibergen. Na het behalen van het MAVO diploma aan de scholengemeenschap Het Assink te Eibergen, werd er in 1997 begonnen met de opleiding Middelbaar Laboratorium Onderwijs, richting Medische Microbiologie en Hematologie aan het ROC te Hengelo. Na het behalen van diploma werd een vervolgopleiding Hoger Laboratorium Onderwijs met richting Medische Biochemie succesvol afgerond in 2005. Tijdens deze studie werd een afstudeerstage gedaan op de afdeling Moleculaire Celbiologie, sectie Virus en Stamcel Biologie van het Leids Universitair Medisch Centrum te Leiden onder de supervisie van dr. M.A.F.V. Gonçalves en dr. A.A.F. de Vries. Vervolgens heeft de auteur vier en een half jaar werkzaam geweest voor dezelfde afdeling als research technician onder de supervisie van dr. M.A.F.V. Gonçalves. Na deze periode werd in 2009 begonnen met het promotie onderzoek op de afdeling Moleculaire Celbiologie, sectie Virus en Stamcel Biologie van de Universiteit Leiden en het Leids Universitair Medisch Centrum onder leiding van Prof. dr. R.C. Hoeben en dr. M.A.F.V. Gonçalves. In 2014 begon de auteur als onderzoeker bij ProQR Therapeutics B.V. te Leiden.

DANKWOORD

Ik ga het niet redden met 1 woord; bedankt. Daarvoor hebben te veel mensen me geholpen in deze lange periode. Daarom ga ik proberen om iedereen te bedanken, maar ik weet nu al dat ik mensen ga vergeten. Lees je dit en sta je er niet tussen; sorry, en alsnog bedankt.

Als eerste wil ik m'n copromotor Manuel bedanken. Ik ben bij je begonnen als student in 2004 en heb veel aan mijn weg van student naar analist en vervolgens dit proefschrift aan je te danken. Ik heb veel van je geleerd, op wetenschappelijk gebied en ook levenslessen. Dit ging met ups en downs, maar aan het einde van de rit zijn er veel meer ups te bedenken. Ik dank je dat ik je "eerste" Ph.D kandidaat mocht zijn.

Mijn promotor Rob Hoeben verdient zeker ook een woord van dank. Je input tijdens de studie en de hulp aan het einde van deze periode waren van grote invloed op het eindresultaat. "Towards the end of my study, I'm expecting a very intriguing question during the defence".

Verder wil ik uiteraard iedereen die in de promotie- en oppositiecommissie wilde plaatsnemen danken. Professoren van der Ploeg, Cathomen, Tijsterman en Staal voor het doornemen van dit proefschrift voordat het besproken mocht worden en Dinko Valerio voor het plaatsnemen in de oppositie commissie.

Vervolgens wil ik mijn paranimfen graag bedanken. Als laatste voor het accepteren en mij helpen verdedigen van dit proefschrift en het vieren ervan, maar als eerste voor een goede vriendschap, zowel tijdens het werk als na het werk. Gijs, je hebt me in het begin van mijn LUMC-carrière werkelijk alles geleerd. Van het perfecte bakje koffie zetten, op tijd beginnen tot het steeds sneller doen van minipreps. Je leerde me "short-cuts" te nemen zonder kwaliteit in te leveren en tegelijkertijd leerde je me wat muziek bij. Je vertrok na een jaar AIO'en naar Rotterdam en ging op zoek naar een virus in de hersenen van MS patiënten maar gelukkig bleven we contact houden. Dirk, als analist kwam je een deel van ons lab claimen en hoewel ik je voorkeur van voetbalclub nog steeds niet kan begrijpen, hebben we in de loop van de tijd een goede vriendschap opgebouwd. Ik kan geen betere paranimfen wensen dan jullie om me te "beschermen" tegen de oppositiecommissie en daarna, om te vieren dat dit hoofdstuk eindelijk afgerond is.

In het loop van de bijna 10 jaar op het LUMC (een jaar als student, 4 jaar als analist en al met al 5 jaar als OIO), heb ik met heel veel mensen samen gewerkt, koffie gedronken, gelachen en af en toe ook nog wetenschap bedreven. Marloes, Marjanka, Michelle, Anna, Vera, John, Shosh, Anabel, Roeland, Jeroen, Ietje, Erik, Jeroen, Jim, Malisa, Harald, Amiet,

D

Martijn, Arnaud, Hester, Laeti, Taco, Jort, Tobias, Daniël, Marjolein, Jin, Chen, Luuk, Joost, Theo, Martijn, Madelon, Lars, Karin, Jort, Anna-Sophia, Niels, Bart, Kim, Gonzo, Amaya, … te veel om op te noemen, maar allemaal bedankt voor alle goede tijd.

Speciaal wil ik Diana, Sanne en Iris bedanken; Robs' angels 2.0, bedankt voor alle praatjes, koffieroddels en camerabeelden van pasgeboren vogeltjes. Sanne en Diana, jullie hebben het Reovirus al getemd, Iris, ik weet zeker dat jij je proefschrift ook af gaat krijgen, de aanhouder wint altijd. Jermain, Jara en Sara; de studenten die ik wat heb mogen bijleren. Bedankt dat jullie mij tegelijkertijd ook veel hebben geleerd, voornamelijk geduld. Steve, nestor van het lab, speciaal voor jou een hele dikke, lange, digitale knuffel. Igno, je bent een soort van mijn opvolger en ik heb met plezier met je samen gewerkt. Je bent slimmer dan je denkt en je gaat het zeker maken in de wetenschap.

Als laatste wil ik Twan bedanken. Je hebt me als student aangenomen en daarna meerdere malen een contract gegeven. Het wijzen op een onbelangrijk detail van een experiment waardoor je het "eureka moment" teniet kon doen hoorde bij je. Evenals het verbeteren van je eigen verbeterde verbeteringen, terwijl je Thomas Dybdahl liedjes neuriede. Halverwege liepen onze wegen de andere kant op maar ik ben blij dat je je plek gevonden hebt bij Cardiologie.

Na mijn tijd bij het LUMC ben ik begin april 2014 begonnen bij ProQR en dat is nu al een anderhalf jaar een soort van achtbaan. Ik wil iedereen bedanken voor nu al een erg mooie tijd en hopelijk duurt dat nog veel langer. Twee mensen wil ik apart bedanken; Tita, je laat me doen wat ik zelf goed denk te doen en ben blij dat ik je als 'baas' heb. Daarnaast wil ik Andre bedanken. Je bent een wandelende CF-wiki die af en toe net iets teveel praat, maar ik heb al zoveel van je geleerd en ik waardeer je visie op wetenschap.

"Elke volleyballer heeft een stelling over volleybal". Deze stelling klopt niet meer, want in de stellingen komt er geen woord volleybal voor. Daarvoor in de plaats en plek in het dankwoord: Heren van Donar H1, het is eindelijk gelukt. Bedankt voor de tijd, trainingen, wedstrijden, 3[de] (4[de], 5[de], 6[de]) setjes, dat er nog velen zullen volgen.

Mijn familie, van beide kanten, bedankt voor alle steun en interesse gedurende al die jaren. Van school, werk en deze studie en zeker ook voor alle grote en kleine dingen. Het was altijd lastig uit te leggen wat ik nou eigenlijk allemaal deed, maar het is eindelijk klaar, het proefschrift is af.

Als laatste wil ik Jolande bedanken. Ik ben niet de makkelijkste, maar toch weet je het al die jaren met me uit te houden. Samen hebben we wat moois opgebouwd, ver van onze families en vrienden en je hebt me in blind vertrouwen en liefde gevolgd naar Leiden. Daarvoor ben ik je nog elke dag dankbaar. Je hebt me altijd gesteund in mijn keuzes, ook al had ik die al gemaakt voordat ik ze aan je had voorgelegd. Nu, 13 jaar na m'n eerste verlegen "hoi" in de lift, wonen we in ons eigen huisje, hebben we 2 geweldige lieve, leuke, eigenwijze, piepende, lachende, stuiterende, vrolijke en knappe meisjes. Dit proefschrift is af, het laatste hoofdstuk is klaar en geschreven maar ik denk en hoop dat ons boek nog lang niet af is en dat er nog veel hoofdstukken bij komen.

Bedankt!
Maarten

D